

10/522037

EPO 24 JAN 2005

**Europäisches
Patentamt****European
Patent Office****Office européen
des brevets**

REC'D 27 NOV 2003

WIPO PCT

Bescheinigung**Certificate****Attestation**

Die angehefteten Unterla-
gen stimmen mit der
ursprünglich eingereichten
Fassung der auf dem näch-
sten Blatt bezeichneten
europäischen Patentanmel-
dung überein.

The attached documents
are exact copies of the
European patent application
described on the following
page, as originally filed.

Les documents fixés à
cette attestation sont
conformes à la version
initialement déposée de
la demande de brevet
européen spécifiée à la
page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

02291871.8

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk**BEST AVAILABLE COPY**



Anmeldung Nr.:
Application no.: 02291871.8
Demande no:

Anmeldetag:
Date of filing: 24.07.02
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Libragen
11 Boulevard Einstein
69100 Villeurbanne
FRANCE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se référer à la description.)

Method for the expression of unknown environmental DNA into adapted host cells

In Anspruch genomene Priorität(en) / Priority(ies) claimed /Priorité(s)
revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

C12N15/00

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of
filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LI LU MC NL PT SE SK TR

Method for the expression of unknown environmental DNA into adapted host cells

Introduction and Background

5 The present invention relates to methods and compositions for nucleic acid production, analysis and cloning. The present invention discloses tools and methods for the production and analysis of libraries of polynucleotides, particularly metagenomic libraries, which can be used to identify novel pathways, novel enzymes and novel metabolites of interest in various areas, including pharmaceutical, cosmetic,
10 agrochemical and/or food industry.

Drug discovery process is based on two main fields, namely combinatorial chemistry and natural products. Combinatorial chemistry has shown its ability to generate huge amounts of molecules, but with limited chemical diversity. At the opposite, natural
15 products have been the most predominant source of structural and molecular diversity. However, the exploitation of this diversity is strongly hampered by their limited access, complex identification and purification processes, as well as by their production.

Microorganisms are known to synthesize a large diversity of natural compounds which
20 are already widely used in therapeutic, agriculture, food and industrial areas. However, this promising approach to the identification of new natural compounds has always been considerably limited by the principal technological bolts of isolating and *in vitro* propagating the huge diversity of bacteria. Most microorganisms living in a natural, complex environment (soil, digestive tract, sea, etc...) have not been cultivated because
25 their optimal living conditions are either unknown or difficult to reproduce. Numbers of scientific publications relate this fact and it is now assumed that less than about 1% of the total bacterial diversity (when all environments are considered together) have been isolated and cultivated (Amann *et al*, 1995).

New approaches have been developed to try to overpass the critical step of isolation, and to access directly to the huge genetic potential established by the microbial adaptation processes through their long evolution. These approaches are called "Metagenomic" because they address a plurality of genomes of a whole bacterial community, without any distinction (metagenome).

Metagenomics involve direct extraction of DNAs from environmental samples and their propagation and expression into a cultivated host cell, typically a bacteria. Metagenomic has been firstly developed for the identification of new bacterial phylum (Pace *et al.* 1997). This use is based on the specific cloning of genes recognized for their interest as phylogenetic markers, such as 16S rDNA genes. Further developments of Metagenomics relate to the detection and cloning of genes coding for proteins with environmental or industrial interest. These first two applications of metagenomic involve a first step of selection of genes (generally using PCR) before cloning. In the case of protein production, the cloning vector used are also expression vectors, i.e., they contain regulatory sequences upstream of the cloning site causing expression of the cloned gene in a given bacterial host strain.

More recent developments of metagenomic consider the total metagenome cloned without any selection and/or identification, to establish random "Metagenomic DNA libraries". This provides an access to the whole genetic potential of bacterial diversity without any "*a priori*" selection. Metagenomic DNA libraries are composed of hundreds of thousands of clones which differ from each other by the environmental DNA fragments which have been cloned. In this respect, large DNA fragments have been cloned (up to 40 Kb), so as to (i) limit the number of clones which have to be analysed and (ii) to be able to recover whole biosynthetic pathways. This last point is of particular interest for bacterial metagenomic libraries since, most of the time, the biosynthetic pathways have been found to be naturally organised in a same cluster of DNA and even in the same operon in bacteria.

Except for the identification and characterisation of bacterial community at the phylogenetic or diversity levels, metagenomic libraries produced in the prior art are gene expression libraries, i.e., the environmental DNA fragments are cloned downstream of a functional promoter, to allow their expression and analysis. In this regard, WO99/45154 and WO96/34112 relate to combinatorial gene expression libraries which comprise a pool of expression constructs where each expression construct contains DNA which is operably associated with one or more regulatory regions that drive expression of genes in an appropriate organisms. Furthermore, the expression constructs used in these methods have a very limited and invariable host range. Similarly, WO 01/40497 relates to the construction and use of expression vectors which can be transferred in one chosen expression bacterial host of the *Streptomyces* genus. All these approaches are, however, very limited since they require the presence of expression signals and confer invariable or very limited host range capabilities. Furthermore, most (if not all) metagenomic DNA libraries are established in *E. coli* which is the most efficient cloning system. However, most environmental DNA are not expressed or functionally active in *E. coli*. In particular, functional analysis in *E. coli* of genes cloned from G+C rich organisms, such as Actinomyces, could be limited by the lack of adequate transcription and translation system. Also, posttranslational modification system in *E. coli* is not operative on heterologous proteins from Actinomicetes and some specific substrates for proteins activity are not present in *E. coli*.

Thus, the exploitation of the huge potential of metagenomics for the discovery of new natural compounds, pathways or genes cannot be achieved with currently existing methods. Alternative technologies and processes must be developed, to allow production of efficient libraries and functional screening in a large variety of host cell species, including *Bacillus* or *Streptomyces*, to take full account of the huge diversity of the environmental DNAs.

Summary of the Invention

The present invention discloses improved tools and methods for the production and analysis of libraries of polynucleotides, particularly metagenomic libraries, which can be used to identify novel pathways, novel enzymes and novel metabolites of interest.

More particularly, the invention now proposes to keep the advantage of high efficient cloning in *E. coli* and to modify the properties of metagenomic libraries, to allow functional analysis of particular selected clones in any appropriate system, thereby making possible the analysis of the huge diversity of metagenomic libraries. According to the invention, polynucleotide libraries can be produced in any convenient cloning system, such as *E. coli*, and then modified, depending on the desired selection or screening system, to adapt host range and/or properties of the library (or a portion thereof).

A particular object of this invention resides more specifically in a method of analysing a library of polynucleotides, said polynucleotides being contained in cloning vectors having a particular host range, the method comprising (i) selecting cloning vectors in the library which contain a polynucleotide having a particular characteristic, (ii) modifying said selected cloning vectors to allow a transfer of said vectors and/or expression of the polynucleotide which they contain into a selected host cell, and (iii) analysing the polynucleotides contained in said modified vectors upon transfer of said modified vectors into said selected host cell.

An other object of this invention is a library of polynucleotides, wherein said library comprises a plurality of environmental DNA fragments cloned into cloning vectors, wherein said environmental DNA fragments contain a common molecular characteristic and wherein said cloning vectors are *E. coli* cloning vectors comprising a target

polynucleotide construct allowing transfer and/or expression of the environmental DNA into a selected host cell distinct from *E. coli*.

5 A further object of this invention is a method of producing modified libraries of polynucleotides, the method comprising selecting a sub-population of clones in a first library, based on the presence or absence of a characteristic of interest, and modifying the properties of said selected clones to allow their functional analysis or expression.

10 The modification in the library or cloning vector is typically obtained by targeted insertion of a polynucleotide construct, preferably using transposable elements, either in vitro or in vivo.

The invention is particularly suited for producing and analysing genetic diversity (metagenomic libraries), to identify and isolate new metabolites, drugs, enzymes, antibiotics, etc.

15

Legend to the Figures

Figure 1 : Map of pP1 vector carrying transposable construct Tn<Apra> fig.1a. Excised transposable construct is 1132 bps in size. It contains two mosaic ends (ME) and a gene conferring resistance to apramycin (Apra) fig.1b.

20

Figure 2 : pPL1 Vector (fig. 2a) that carries the conjugative transposable construct Tn<Apra-oriT> (fig. 2b). The nucleotide sequence of the transposable construct contains an origin of transfer (oriT) and a gene conferring resistance to apramycin (Apra). Direction of DNA transfer at oriT is shown by an arrow.

25

Figure 3 : Transposable construct Tn<Apra-oriT-att-int> (fig. 3a) on vector pPAOI6 (fig. 3b). The transposable construct contains Φ C31 integrase gene and attachment DNA

sequence for site specific integration, origin of transfer and gene for selection. The orientation of genes and direction of DNA transfer are marked by arrows.

Figure 4 : This figure shows any target DNA suitable for insertion of transposable construct (fig. 4a). Insertion of conjugative and site specific integrative transposable construct Tn<Apra-oriT-int> is shown on fig. 4 b and c. Insertion of transposable construct into selective gene marker carried on original vector is shown on fig. 4b. In this event, the cloned insert is intact and can be transferred to heterologous host. Insertion of transposable construct into cloned DNA insert results in gene inactivation (fig. 4c).

Figure 5 : Complete annotated DNA sequences of clones FS3-124 (a, SEQ ID NO: 1) and FS3-135 (b, SEQ ID NO: 2).

Figure 6 : Morphological differences between *Streptomyces* transconjugant (assay). Conjugations have been performed with FS3-124 modified with transposable construct pPAOI6; (control) conjugation have been performed with pPAOI6.

Detailed Description of the Invention

20

The invention provides novel strategies, methods and products for generating and analysing combinatorial gene libraries. As indicated above, the invention discloses, particularly, methods of analysing libraries of polynucleotides, said polynucleotides being contained in cloning vectors having a particular host range, the methods comprising (i) selecting cloning vectors in the library which contain a polynucleotide having a particular characteristic, (ii) modifying said selected cloning vectors to allow a transfer of said vectors and/or expression of the polynucleotide which they contain into a selected host cell, and (iii) analysing the polynucleotides contained in said modified

vectors upon transfer of said modified vectors into said selected host cell, such as by genetic, biochemical, chemical or phenotypical approaches.

Library of polynucleotides

5

The term "library of polynucleotides" designates a complex composition comprising a plurality of polynucleotides, of various origins and structure. Typically, the library comprises a plurality of unknown polynucleotides, i.e., of polynucleotides whose sequence and/or source and/or activity is not known or characterized. In addition to such
10 unknown (or uncharacterised) polynucleotides, the library may further include known sequences or polynucleotides. Typically, the library comprises more than 20 distinct polynucleotides, more preferably at least 50, typically at least 100, 500 or 1000. The complexity of the libraries may vary. In particular, libraries may contain more than 5000, 10 000 or 100 000 polynucleotides, of various origin, source, size, etc. Furthermore, the
15 polynucleotides are generally cloned into cloning vectors, allowing their maintenance and propagation in suitable host cells, typically in *E. coli*. The polynucleotides in the library may be in the form of a mixture or separated from each other, in all or in part. It should be understood that some or each polynucleotide in the library may be present in various copy numbers.

20

The polynucleotides in the libraries are more preferably obtained or cloned from complex sources of nucleic acids, most preferably from environmental samples. Such libraries are also termed "metagenomic libraries" since they contain nucleic acids derived from whole genomes of complex mixtures of microorganisms.

25

The term environmental sample designates, broadly, any sample containing (a plurality of) uncharacterized (micro)organisms, particularly uncultivated (or non-cultivable) microorganisms. The sample may be obtained or derived from specific organisms, natural environments or from artificial or specifically created environments (e.g.,

industrial effluents, etc). An uncultivated (or non-cultivable) microorganism is a microorganism that has not been purposely cultured and expanded in isolated form. The sample may be obtained or derived from soil, water, mud, vegetal extract, wood, biological material, marine or estuarine sediment, industrial effluents, gas, mineral
 5 extracts, sand, natural excrements, meteorits etc. The sample may be collected from various regions or conditions, such as tropical regions, deserts, volcanic regions, forests, farms, industrial areas, household, etc.

Environmental samples usually contain various species of (uncharacterized,
 10 uncultivated) microorganisms, such as terrestrial microorganisms, marine microorganisms, salt water microorganisms, freshwater microorganisms, etc. Species of such environmental microorganisms include autotrophe or heterotrophe organisms, Eubacteria, archaebacteria, algae, protozoa, fungi, viruses, phages, parasites, etc. The microorganisms may include extremophile organisms, such as thermophiles,
 15 psychrophiles, psychrotophes, acidophiles, halophiles, etc. More specific examples of environmental bacteria includes actinomycetes, eubacteriaes and mycobacteriaes, examples of fungi include phycomycetes, ascomycetes and basidiomycetes, etc. Other organisms include yeasts (saccharomyces, kluyveromyces, etc.) plant cells (algae, lichens, etc), corals, etc. for instance. The sample may comprise various species of such
 20 donor (uncultivated) microorganisms, as well as various amounts thereof. The environmental sample may contain, in addition, known and/or cultivable microorganisms (e.g., prokaryotic or eukaryotic), as well as nucleic acids and organic materials. The sample may also contain mammalian cells or animal cells or insect cells (arising from larvae, feces, etc.).

25

It should be understood that the present invention is not limited to any specific type of sample or environmental microorganism, but can be used to produce diversity, create nucleic acid libraries, etc., from any environmental sample comprising uncultivated microorganisms. The sample may be wet, soluble, dry, in the form of a suspension,

paste, powder, solid, etc. Preferably, the sample is dry or in solid or semi-solid state (e.g., paste, powder, mud, gel, etc.). The sample may be treated prior to nucleic acid extraction, for instance by washings, filtrating, centrifuging, diluting, drying, etc.

- 5 The term "environmental DNA" designates any DNA fragment or collection obtained from an environmental sample. Nucleic acids may be extracted / isolated from the sample according to various techniques, such as those described in WO01/81357, in WO01/40497, in Handelsman *et al.* (Chemistry & Biology 5(10), 1998, R245), Rondon *et al.* (Tibtech 17, 1999, 403 ; Applied and Environm. Microbiol. 66, 2000, 2541), Miller
10 et al (Applied and Environm. Microbiol. 65, 1999, 4715) or Frostegard *et al.* (Applied and Environm. Microbiol., 65, 1999, 5409).

- In a particular embodiment of the above method, the library comprises a plurality of environmental DNA fragments. The library may also comprise other types of nucleic
15 acids, such as environmental RNAs, for instance.

- The polynucleotides have a size which is typically comprised between 10 and 100 kb, more preferably between 20 and 80 kb, typically between 30 and 80 kb. Although not mandatory, it is preferred that the polynucleotide fragments in the library all have similar
20 size, to produce homogenous libraries.

Cloning Vectors

- As indicated, the polynucleotides are contained or cloned into cloning vectors. These
25 vectors may be of various types, including plasmids, cosmids, fosmids, episomes, artificial chromosomes, phages, viral vectors, etc. In most preferred embodiments, the cloning vectors are selected from plasmids, cosmids, phages (e.g., P1-derivatives) and BACs, even more preferably from cosmids, P1 derivatives and BACs. By using cosmids or P1 derivatives, it is possible to generate homogenous libraries, since these vectors

essentially accommodate polynucleotides having a size of approximately 40 kb and 80 Kb, respectively. Furthermore, since the invention provides that the vectors are modified after the initial cloning step, the cloning capacity of the vectors is maximized and inserts of as much as 40 and 80 kbs in length can be cloned into cosmids and P1 derivatives.

5

As indicated the cloning vector has a particular host range, i.e., the ability to replicate in a particular type of host cell. Typically, the host is a bacteria, more preferably an *E. coli* strain. Indeed, *E. coli* is so far the most convenient host cell for performing recombinant technologies. The advantage of the present invention is that the starting library can be produced in any suitable host system of choice, since the properties of the libraries will be adapted later during the process.

10

Cloning vectors generally comprise the polynucleotide insert and genetic elements necessary and sufficient for maintenance into a competent host cell. They typically contain, in addition to the polynucleotide insert, an origin of replication functional in a selected host cell as well as a marker gene for selection and screening. The cloning vector may comprise additional elements, such as promoter regions, for instance. Although cloning vectors may replicate in several different host cells, they are usually adapted to a particular host cell type and not suitable or efficient for replication or maintenance in other cell types.

15

20

In a preferred embodiment, the cloning vectors of the library are *E. coli* cloning vectors, preferably cosmids, BAC or P1 derived vectors. *E. coli* cloning vectors may carry an origin of replication derived from naturally-occurring plasmids, such as ColE1, pACYC and p15A, for instance. Many *E. coli* cloning vectors are commercially available and/or can be constructed using available regulatory sequences.

25

Screening of the cloning vectors

In step i) of the method, a first selection or screen is performed on the polynucleotide library. The screen is performed so as to identify or select clones having (or lacking) a particular, common characteristic. The selection may be carried out according to various techniques, such as by molecular screening, protein expression, functional screening, etc.

- 5 A preferred selection is performed by molecular screening. Molecular screening designates any method of identification of molecular or structural characteristics in a polynucleotide sequence. This can be made by a variety of techniques which are known per se, such as hybridisation, amplification, sequencing, etc. Preferably, molecular screening comprises the selection of clones in a library which contain, in their sequence,
- 10 a particular sequence or region or motif, said sequence or region or motif being characteristic of a particular type of activity or gene (enzyme, biosynthetic pathways, etc.).

- In a first variant, the selection is made by contacting the cloning vectors in the library
- 15 with a particular nucleic acid probe (or set of probes) containing a sequence which is characteristic of a selected activity or function (a consensus sequence, a particular motif, etc.). The cloning vectors in the library which hybridise to the probe (or set of probes) are then selected.

- 20 In a second variant, the selection is made by contacting the cloning vectors in the library with a particular pair of nucleic acid primers specific for a sequence which is characteristic of a selected activity or function (a consensus sequence, a particular motif, etc.), and a PCR amplification reaction is performed. The cloning vectors in the library which lead to a positive amplification product are then selected.

25

In this regard, the present application provides new primers designed in conserved motives of the β -keto acyl synthase gene, which are particularly useful for screening polynucleotides containing putative PKS genes or domains. These primers have the following degenerated sequence :

Sense primer : 5'- GGSCCSKCSSTSDCSRTSGAYACSGC -3' (SEQ ID NO: 3)

Antisense primer : 5'- GCBBSRYYTCDATSGGRTCSCC -3' (SEQ ID NO: 4)

wherein :

R is A or G

5 S is G or C

Y is C or T

K is G or T

D is A or G or T, and

B is C or G or T.

10

A particular object of this invention is a polynucleotide primer having one of the above sequences, typically a mixture of different polynucleotide primers having a sequence corresponding to one of the above degenerated sequences. A particular object of this invention also resides in a pair of primers each having one of the above sequences.

15

Once particular clones have been selected, the analysis of their polynucleotides needs to be confirmed and/or validated, and/or their polynucleotides can be used to produce novel compounds or metabolites.

20 The invention now enables such further analysis and uses, by allowing a modification of the cloning vectors that is specific and adaptable by the skilled person, depending on the activity which is sought. In particular, it is possible to confer properties such as specific expression or a novel, specific host range to the selected cloning vectors, to assess their activity, as disclosed below.

25

Modification of the cloning vectors

After high efficiency cloning using most convenient cloning vectors such as BACs or cosmids propagated into *E. coli*, and after the identification, selection and/or

characterisation of cloned DNA fragments, the invention now allows to modify specifically the cloning vectors to transfer, maintain, express and/or over-express the selected polynucleotides into any selected host expression system, which is suitable to assess the selected activity or property. Such selected hosts may be native or
5 heterologous host cells, and include, but are not limited to, for example *Streptomyces*, *Nocardia*, *Bacillus*, fungi, yeasts, etc.

The selected cloning vectors of the library may be modified according to various techniques. The modification is typically a genetic modification, comprising the
10 introduction of particular genetic sequences into the structure of the cloning vector, in addition to or in replacement of sequences contained in said vector. It is highly preferred to use specific or targeted (or oriented) techniques to improve the efficacy of the method. By "specific" is meant that the modification occurs at a pre-determined location in the cloning vector, through site-specific mechanisms. By "targeted" is meant that the
15 modification occurs in a controlled way, so as not to alter the polynucleotide insert contained in the vector in a non-desirable way.

In a preferred embodiment, the selected vectors are modified by insertion, into the vector, of a target polynucleotide construct which contains genetic elements conferring
20 the selected property(ies) to the cloning vector.

The target polynucleotide construct typically comprises the genetic elements necessary to transfer, propagate, maintain, express or overexpress the cloned polynucleotide into a chosen (bacterial) host expression system. Said genetic elements may include particular
25 origin(s) of replication, particular origin(s) of transfer, particular integrase(s), transcriptional promoter(s) or silencer(s), either alone or in combination(s).

In a first, preferred variant, the target polynucleotide construct comprises a genetic element allowing transfer of the vector into a selected host cell.

Natural DNA transfer mechanisms between donor and recipient strains is known under the term conjugation or conjugative transfer. Conjugative transfer can occur between different strains of the same species as well as between strains of different species. Many naturally occurring plasmids carry so called *tra* genes, which are involved in and mediate conjugative transfer. The DNA transfer starts at specific DNA structures, known as an origin of transfer or "ori T". The presence of such an oriT in a vector allows said vector to be transferred into a desired host cell.

In a particular, preferred embodiment, the target polynucleotide construct comprises an origin of transfer functional in the selected host cell, more particularly in a *Streptomyces* or *Bacillus* host cell.

The structure of various oriT has been reported in the art (Guiney *et al.* 1983; Zechner *et al.* 2000). In a specific embodiment, the origin of transfer is selected (or derived) from RP4, pTiC58, F, RSF1010, ColE1 and R6K(α).

A specific example of an oriT which can be used in the present invention derives from plasmid RP4 and has or comprises all or a functional part of the following sequence (SEQ ID NO: 5) :

gatctGTGATGTACTTCACCAGCTCCGCGAAGTCGCTCTTCTTGATTGGAGCGCATGGG
 GACGTGCTTGGCAATCACGCGCACCCCCCGGCCGTTTTAGCGGCTAAAAAAGTCATG
 GCTCTGCCCTCGGGCGGACCACGCCATCATGACCTTGCCAAGCTCGTCCTGCTTCT
 25 CTTGATCTTCGCCAGCAGGGCGAGGATCGTGGCATCACCGAACCGCGCCGTGCGC
 GGGTCGTTCGGTGAGCCAGAGTTTCAGCAGGCCGCCAGGCGGCCAGGTGCGCCATT
 GATGCGGGCCAGCTCGCGGACGTGCTCATAGTCCACGACGCCCGTGATTTTGTAGCC
 CTGGCCGACGGCCAGCAGGTAGGCCGACAGGCTCATGCCGGCCGCCGCGCCTTTTC
 CTCAATCGCTCTTCGTTTCGTCTGGAAGGCAGTACACCTTGATAGGTGGGCTGCCCTT
 30 CCTGGTTGGCTTGGTTTTCATCAGCCATCCGCTTGCCCTCATCTGTTACGCCGGCGGTA
 GCCGGCCAGCCTCGCAGAGCAGGATTCCCGTTGAGCACCGCCAGGTGCGAATAAGG
 GACAGTGAAGAAGGAACACCCGCTCGCGGGTGGGCCTACTTCACCTATCCTGCCCG
 GCTGACGCCGTTGGATACACCAAGGAAAGTCTACACGAACCCTTTGGCAAAATCCT

GTATATCGTGCGAAAAAGGATGGATATACCGAAAAAATCGCTATAATGACCCCGAA
GCAGGGTTATGCAGCGGAAAAGATCCGTCGGATCT

The term "functional part" designates any fragment or variants of the above sequence
5 which retain the capacity to cause conjugative transfer. Such fragments typically
comprise at least 80%, preferably at least 85% or 90% of the above sequence. Variants
may include one or several mutations, substitutions, deletions or additions of one or
several bases.

10 In an other particular variant, the target polynucleotide construct comprises a genetic
element allowing integration of the vector into the genome of the selected host cell.

A donor DNA is permanently or stably maintained and expressed in a selected recipient
cell if it is integrated into the recipient cell's genome or if it contains elements that allow
15 autonomous replication in said cell.

Integration is a preferred way of ensuring stable expression. Integration can be obtained
by physical recombination. Recombination can be homologous, e.g., between two
homologous DNA sequences, or illegitimate, where recombination occurs between two
20 non-homologous DNAs. As a particular example, integration of donor DNA into the
chromosome of the recipient can be mediated by host recombination repair system or by
site-specific recombination. Another well-studied process that can transfer and integrate
genes is transduction by bacterial viruses, such as λ and ϕ C31. In a phage-infected
bacterial cell, fragments of the host DNA are occasionally packaged into phage particles
25 and can then be transferred to a recipient cell. Integration into the recipient cell's genome
is caused by an integrase.

In a specific embodiment, the target polynucleotide construct comprises a nucleic acid
encoding an integrase functional in the selected host cell. More preferably, the integrase
30 is selected from λ and ϕ C31 integrases. In a specific embodiment, the polynucleotide

construct comprises a nucleic acid encoding an integrase having or comprising all or a functional part of the following sequence of the ϕ C31 integrase (SEQ ID NO: 6) :

```

5  AGATCTCCCGTACTGACGGACACACCGAAGCCCCGGCGGCAACCCTCAGCGGATGC
   CCCGGGGGCTTCACGTTTTCCCAGGTGAGAAGCGGTTTTTCGGGAGTAGTGCCCCAACT
   GGGGTAACCTTTGAGTTCTCTCAGTTGGGGGCGTAGGGTCGCCGACATGACACAAG
   GGGTTGTGACCGGGGTGGACACGTACGCGGGTGCTTACGACCGTCAGTCGCGCGAG
   CGCGAGAATTCGAGCGCAGCAAGCCCAGCGACACAGCGTAGCGCCAACGAAGACA
   AGGCGGCCGACCTTCAGCGCGAAGTCGAGCGCGACGGGGGCGGGTTCAGGTTCTGTC
10  GGGCATTTCAGCGAAGCGCCGGGCACGTGCGCGTTCGGGACGGCGGAGCGCCCGGA
   GTTCGAACGCATCCTGAACGAATGCCGCGCCGGGCGGCTCAACATGATCATTGTCTA
   TGACGTGTGCGGCTTCTCGCGCCTGAAGGTCATGGACGCGATTCCGATTGTCTCGGA
   ATTGCTCGCCCTGGGCGTGACGATTGTTTCCACTCAGGAAGGCGTCTTCCGGCAGGG
   AAACGTGTCATGGACCTGATTACCTGATTATGCGGCTCGACGCGTCGCACAAAGAATC
15  TTCGCTGAAGTCGGCGAAGATTCTCGACACGAAGAACCTTCAGCGCGAATTGGGCG
   GGTACGTGCGCGGGAAGGCGCCTTACGGCTTCGAGCTTGTTTCGGAGACGAAGGAG
   ATCACGCGCAACGGCCGAATGGTCAATGTCGTCAACAAGCTTGCGCACTCGACC
   ACTCCCCTTACCGGACCCTTCGAGTTCGAGCCCGACGTAATCCGGTGGTGGTGGCGT
   GAGATCAAGACGCACAAACACCTTCCCTTCAAGCCGGGCAGTCAAGCCGCCATTCA
20  CCCGGGCAGCATCACGGGGCTTTGTAAGCGCATGGACGCTGACGCCGTGCCGACCC
   GGGGCGAGACGATTGGGAAGAAGACCGCTTCAAGCGCCTGGGACCCGGCAACCGTT
   ATGCGAATCCTTCGGGACCCGCGTATTGCGGGCTTCGCCGCTGAGGTGATCTACAAG
   AAGAAGCCGGACGGCACGCGGACCACGAAGATTGAGGGTTACCGCATTACGCGCGA
   CCCGATCACGCTCCGGCCGGTCGAGCTTGATTGCGGACCGATCATCGAGCCCGCTGA
25  GTGGTATGAGCTTCAGGCGTGTTGGACGGCAGGGGGCGCGGCAAGGGGGCTTTCCC
   GGGGGCAAGCCATTCTGTCCGCCATGGACAAGCTGTACTGCGAGTGTGGCGCCGTC
   ATGACTTCGAAGCGCGGGGAAGAATCGATCAAGGACTCTTACCGCTGCCGTGCCG
   GAAGGTGGTCGACCCGTCCGCACCTGGGCAGCACGAAGGCACGTGCAACGTCAGCA
   TGGCGGCACTCGACAAGTTTCGTTGCGGAACGCATCTTCAACAAGATCAGGCACGCC
30  GAAGGCGACGAAGAGACGTTGGCGCTTCTGTGGGAAGCCGCCCGACGCTTCGGCAA
   GCTCACTGAGGCGCCTGAGAAGAGCGGCGAACGGGCGAACCTTGTTGCGGAGCGCG
   CCGACGCCCTGAACGCCCTTGAAGAGCTGTACGAAGACCGCGCGGCAGGCGCGTAC
   GACGGACCCGTTGGCAGGAAGCACTTCCGGAAGCAACAGGCAGCGCTGACGCTCCG
   GCAGCAAGGGGCGGAAGAGCGGCTTGCCGAACCTTGAAGCCGCCGAAGCCCCGAAG
35  CTTCCCCTTGACCAATGGTTCCCCGAAGACGCGGACGCTGACCCGACCGGCCCTAAG
   TCGTGGTGGGGGCGCGCGTCAGTAGACGACAAGCGCGTGTTCTGTCGGGCTCTTCGTA
   GACAAGATCGTTGTACGAAGTCGACTACGGGCAGGGGGCAGGGAACGCCCATCGA
   GAAGCGCGCTTCGATCACGTGGGCGAAGCCGCCGACCGACGACGACGAAGACGACG
   CCCAGGACGGCACGGAAGACGTAGCGGCGTAGCGAGACACCCG
40

```

In a more preferred variant, the target polynucleotide construct comprises genetic elements allowing transfer of the cloning vector into the selected host cell and integration of the cloning vector or a portion thereof into the genome of the selected host

cell. Most preferred polynucleotide constructs comprises an oriT and a nucleic acid encoding an integrase.

5 In an other variant, the target polynucleotide construct comprises an origin of replication specific for or functional in the selected host cell. The origin of replication may be selected (or derived), for instance, from pAM β 1, pSa, 2 μ m circle, pSam2, pSG1, pIJ101, SCP2, pA387 and artificial chromosomes.

10 In an other variant, the target polynucleotide construct comprises a transcriptional promoter functional in the selected host cell. As indicated above, in a particular variant, the invention allows to modify the cloning vector to enable expression or over-expression of the cloned polynucleotides in the selected host. The expression of genes is driven mainly by transcriptional promoters, which initiate gene transcription. The type of promoter to be used in the present invention can be selected by the skilled person,
15 depending on the selected host cell and type of expression needed. Promoters may be ubiquitous or cell-specific, regulated or constitutive, weak or strong. They may be of various origins, including promoters isolated from viruses, phages, plant cells, bacterial genes, mammalian genes, etc., or they may be artificial or chimeric. Typical examples of promoters include T7, T4, LacZ, trp, ara, SV40, tac, λ PL, GAL, AOX, hsp-70, etc.

20 The target polynucleotide construct is typically a DNA molecule, although RNAs may also be used as starting material. It is typically a double-stranded DNA. The target polynucleotide construct may be produced by conventional recombinant DNA techniques, including DNA synthesis, cloning, ligation, restriction digestion, etc. and a
25 combination thereof.

The target polynucleotide construct is preferably engineered so as to be inserted in a region of the vector distinct from the polynucleotide. Indeed, it is important that the integrity of the polynucleotides is preserved. Directed insertion may be accomplished in

a variety of ways, including site-specific insertion using particular enzymatic systems (Cre/Lox, FLP, etc.), homologous recombination with particular target sequences present in the vector, or by the use of transposons or transposable elements and appropriate selection means.

5

In a particular, preferred embodiment, the target polynucleotide construct is contained in a transposable nucleic acid construct. Indeed, in a preferred variant, the methods of the present invention use transposable elements to alter the properties of the cloning vectors, and allow their transfer, maintenance, expression or over-expression in a selected host
10 cell.

Transposable nucleic acid constructs are derived from transposons, which are genetic elements capable of moving from one genetic loci to another. Two main classes of transposons have been identified in bacteria. The most simple transposons comprise an
15 insertion sequence that carries only elements of transposition. These elements are two inverted DNA repeats and a gene that codes for a protein called transposase. The transposase catalyses the excision and integration of the transposon. It has been shown that the excision and integration reaction can be catalysed in trans by a transposase, which can be provided in vivo or in vitro in purified form or expressed from a different
20 construct. More complex transposons carry more insertion sequences and additional genes that are not involved in transposition.

Transposable nucleic acid constructs of this invention thus typically comprise, flanked by two inverted repeats, the target polynucleotide construct and, more preferably, a
25 marker gene. In the presence of a transposase, these transposable nucleic acid constructs can integrate into a cloning vector in vivo or in vitro, thereby providing for targeted polynucleotide insertion. Alternatively, such nucleic acid constructs can be used for targeted integration, in the absence of a transposase, in particular strains such as hypermutator strains. Such transposable nucleic acid constructs also represent a

particular object of the present application. In this regard, in a more preferred embodiment, the invention also relates to a transposable nucleic acid construct, wherein said construct comprises an origin of transfer flanked by two inverted repeats. Specific examples of such construct are transposons pPL1 and pPAOI6, as disclosed in the experimental section. The transposable nucleic acid construct may further comprise an integrase gene and/or a marker gene.

The inverted repeat nucleic acid sequences may be derived from the sequence of various transposons, or artificially created. In particular, transposable elements can be generated using inverted repeats obtained from transposons or transposable elements such as Tn5, Tn21, miniTn5, T7, T10, Tn917, miniTn400, etc. Preferably, the sequences derive from transposon Tn5. In a specific embodiment, they comprise all or a functional part of the following sequences :

- 15 - left arm of pPAOI6 transposon (SEQ ID NO: 7)
CTGTCTCTTATACACATCTCAACCATCATCGATGAATTTTCTCGGGTGTTCCTCGCATA
TTGGCTCGAATTCGAGCTCGGTACCC
- right arm of transposon pPAOI6 (SEQ ID NO: 8)
20 GATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGCCAACGACTACGCACTAGCCA
ACAAGAGCTTCAGGGTTGAGATGTGTATAAGAGACAG

The marker gene may be any nucleic acid encoding a molecule whose presence in a cell can be detected or visualized. Typical marker genes encode proteins conferring resistance to antibiotics, such as apramycin, chloramphenicol, ampicillin, kanamycin, spectinomycin, thiostrepton, etc. Other types of markers confer auxotrophy or produce a label (e.g., galactosidase, GFP, luciferase, etc).

In a specific embodiment, the cloning vector in the library comprises a first marker gene and the modification step ii) comprises:

- . contacting in vitro, in the presence of a transposase, the selected cloning vectors with a transposon comprising, flanked by two inverted repeats, the target polynucleotide construct and a second marker gene distinct from the first marker gene, and
5 . selecting the cloning vectors which have acquired the second marker gene and which have lost the first marker gene.

The double selection ensures that the target polynucleotide construct has been inserted at a site within the marker gene present in the cloning vector, i.e., outside of the polynucleotide insert.

10

- It should be understood that the modification may be accomplished in various other ways, particularly by incorporating a sequence coding for the transposase directly into the transposable element or into another expression unit. The presence and expression of the transposase can be regulated by inductive promoter or termosensitive replicative
15 units. Also, transposition can be carried out by *in vitro* process.

Analysis of the polynucleotides

- In step (iii) of the process, the polynucleotides may be analysed by various methods,
20 including by genetic, biochemical, chemical or phenotypical approaches, which are well known per se in the art. Analysis occurs upon transfer and, optionally, expression of the polynucleotides into the selected host cell.

- In this regard, the modified cloning vectors can be transferred into the selected host cell
25 by a variety of techniques known in the art, including by transformation, electroporation, transfection, protoplast fusion, conjugative transfer, etc. In a preferred embodiment, the target polynucleotide construct comprises an oriT and the modified vectors are transferred into the selected host cells by conjugative transfer. In this embodiment, the

cloning vector and the selected host cells are co-cultivated and the recombinant host cells are selected and isolated.

The selected host cell may be any type of cell or microorganism, including, without
5 limitation, *Streptomyces*, *E. coli*, *Salmonella*, *Bacillus*, *Yeast*, *fungi*, etc.

One of the objectives of the invention is to be able to analyse environmental DNAs of unknown cellular origin into different host expression systems. In order to analyse the potentiality of the DNAs at the transcription and/or translation levels and to have much
10 more probabilities to have a DNA expression, it is important to have the possibility to test different host expression systems.

Insertion of foreign DNA into host expression systems like *Streptomyces* can produce, for instance, an increase in doubling time, morphological modifications, pigments
15 production, etc. which can be related either directly to the expression of foreign DNA or by combinatorial biology of the foreign DNA and the biology of the expression host systems. The new characters can be analysed by all techniques known in the art such genetic, biochemical, chemical, phenotypic approaches etc...

20 In a preferred, specific embodiment, the invention relates to a method for the identification or cloning of polynucleotides encoding a selected phenotype, the method comprising (i) cloning environmental DNA fragments into *E. coli* cloning vectors to produce a metagenomic library, (ii) identifying or selecting cloning vectors in said library which contain DNA fragments having a particular characteristic of interest, (iii)
25 modifying the identified or selected cloning vectors into shuttle or expression vectors for transfer or expression in a selected host cell, (iv) transferring or expressing the modified cloning vectors into said selected host cell and (v) identifying or cloning the DNA fragments contained in said modified cloning vectors which encode said selected phenotype in said selected host cell.

By applying the above method, new polynucleotide sequences have been identified, cloned and characterized, which produce new phenotypes in bacteria. These polynucleotides contain the sequence of PKS genes and encode polypeptides involved in biosynthetic pathways. The sequence of these polynucleotides is provided in Figure 5. The invention also relates to any polynucleotide sequence comprising all or part of these sequences (i.e., SEQ ID NOs: 1 or 2), their complementary strand, or a functional variant thereof. A part of the above sequences includes, preferably, at least 20 consecutive bases, more preferably at least 50 consecutive bases thereof, even more preferably a coding sequence (e.g., an ORF). In this respect, SEQ ID NOs: 1 and 2 comprise several novel open reading frames encoding novel polypeptides involved in biosynthetic pathways. These coding sequences are identified in Figures 5a and 5b.

In a specific embodiment, the invention relates to a polynucleotide sequence comprising a sequence selected from nucleotides 76 - 1134 ; 1096 - 2430 ; 1178 - 1624 ; 2506 - 3567 ; 2906 - 4222 ; 4092 - 5321 ; 6337 - 8502 ; 8181 - 9530 ; 9531 - 10721 ; 10504 - 11274 ; 12874 - 13689 ; 14195 - 15976 ; 15427 - 16512 ; 15579 - 16253 ; 16505 - 17656 ; 17657 - 18697 ; 18615 - 19304 ; 19301 - 20596 ; 20535 - 21476 ; 22025 - 22951 ; 23155 - 26523 ; 26409 - 34433 ; 34418 - 37500 and 35359 - 37500 of SEQ ID NO: 1 (figure 5a) or a complementary strand thereof.

In an other specific embodiment, the invention relates to a polynucleotide sequence comprising a sequence selected from nucleotides 3 - 914, 924 - 2168 ; 2207 - 3190 ; 3373 - 4455 ; 4546 - 4959 ; 5176 - 6192 ; 6331 - 14043 ; 14275 - 15408 ; 15436 - 16245 ; 16287 - 17384 ; 17427 - 18158 ; 18248 - 18847 ; 18952 - 20346 ; 20442 - 21167 ; 21164 - 24301 ; 24351 - 27023 ; 27806 - 29686 ; 29535 - 30872 ; 30848 - 32647 ; 32574 - 35555 ; 35533 - 36598 and 36516 - 37400 of SEQ ID NO: 2 (figure 5b) or a complementary strand thereof.

The invention also relates to any vector comprising these polynucleotide sequences. These sequences may be DNA or RNA, preferably DNA, even more preferably double-stranded DNA. The invention also relates to a polypeptide encoded by a polynucleotide sequence as defined above.

5

An other object of this invention resides in a library of polynucleotides, wherein said library comprises a plurality of environmental DNA fragments cloned into cloning vectors, wherein said environmental DNA fragments contain a common molecular characteristic and wherein said cloning vectors are *E. coli* cloning vectors comprising a
10 target polynucleotide construct allowing transfer or expression of the environmental DNA into a selected host cell distinct from *E. coli*.

The sub-DNA-libraries should have either desired genetic characteristics based on high or low GC content, DNA encoded for a desired enzymatic activity, part or full
15 biosynthetic pathways for metabolites etc., or specific origin such as soil fractions, animal organs, sub fraction of a microorganism community etc. The invention also allows to produce conjugative vector with desired characteristics in accordance with the characteristics of the pre-identified sub-DNA-libraries and functional analysis of mutants in heterologous hosts. It can also be used, without limitation, for the production of
20 mutants by random or directed mutagenesis, for DNA sequencing, genes or biosynthetic pathways knock-out by deletion or insertion or to confer transfer capabilities for expression, co-expression, over-expression or modification of biosynthetic pathways.

Further aspects and advantages of the invention will be disclosed in the following
25 examples, which should be regarded as illustrative and not limiting the scope of this application.

Experimental Section

In this work, we constructed a fosmid library in *E. coli* from DNA prepared directly from soil. We screened the library for presence of biosynthetic pathways. We developed genetic tools for functional genomics that allow gene inactivation and horizontal gene transfer from *E. coli* to *Streptomyces*.

5

The cloning vectors in the library contain ColE1 replicon for propagation in *E. coli*. Transposable elements based on Tn 5 transposon were produced and used for *in vitro* (or *in vivo*) modification of selected cloning vectors. Integrated transposable elements contain gene for resistance to apramicyne. Conjugative derivatives were constructed by incorporating origin of transfer from plasmid RP4. A conjugative and site specific integrative transposon was also constructed comprising the integrase gene from ϕ C31 phage, including attP attachment site. Conjugal transfer was demonstrated from an appropriate *E. coli* donor cell to another *E. coli* or *Streptomyces lividans* recipient cell. Constructed transposon was tested for inactivation of the genes cloned into fosmids.

15 Obtained mutants can be used for direct sequencing by adequate primers and transferred by conjugation into *Streptomyces lividans*. Transposable elements thus represent very useful tools for functional analysis of a large DNA libraries cloned into BAC, PAC fosmids or other cloning vectors in which cloned inserts must be transferred into heterologous host.

20 **Materials and methods**

Bacterial strains, plasmids and growth conditions.

E. coli DH10B (F- mcrA delta(mrr-hsdRMS-mcrBC) phi80dlacZ deltaM15 delta lacX74 deoR recA1 endA1 araD139 delta (ara, leu)7697 galU galK lambda- rpsL nupG), strain (Epicentre) was used for fosmid and plasmid transformation and DNA amplification.

25 Unless specifically described, all DNA manipulations were performed according to Sambrook, J., *et al*, (1989).

Soil DNA extraction and DNA libraries construction.

Total bacterial community DNA was extracted and large DNA libraries have been constructed into fosmids according to the method described in WO 01/81357.

Fosmid DNA extraction and Purification

5

Fosmids DNA containing soil library were extracted from pools of 96 clones. Culture of recombinant clones were performed in Deep-Well 96 and 48, respectively in 1 ml and 2 ml of LB media containing 12,5 $\mu\text{g ml}^{-1}$ of chloramphenicol. Cultures were grown at 37°C with shaking at 250 RPM during 22 hours. DNA extraction was done by using the

10 Nucleobond PC100 extraction kit (Macherey Nagel).

PCR screening for the detection of PKS genes

Primers design

Degenerate PCR primers sets were designed to specifically amplify PKS nucleic
15 sequences. Multiple sequence alignment of PKS domains revealed highly conserved motives, in particular in the β -keto acyl synthase domain. Primers Lib1F and Lib2R were designed in conserved motives of the β -keto acyl synthase gene. Lib1F (sense primer, 5'- GGSCCSKCSSTSDCSRTSGAYACSGC -3') and Lib2R (antisense primer, 5'- GCBBSRYTCDATSGGRTCCSCC -3') were deduced from β -keto acyl synthase
20 peptide sequences GP(AS)(LV)(AST)(IV)DTAC and GDPIE(TVA)(RAQ)A, respectively. The specific fragment amplified with Lib1F / Lib2R was approximately about 465 bp (corresponding to 155 amino acids). Specificity and efficiency of the PCR systems were validated by testing on positive DNA controls (i.e. genomic DNA from type I PKS producing strain such as *Bacillus subtilis*, *Streptomyces lividans*,
25 *Streptomyces ambofaciens* and *Ralstonia solanocearum*) and negative DNA controls (genomic DNA from strains which are known to do not contain PKS genes). Furthermore, DNA extracted from soil samples were tested to calibrate PCR techniques.

PCR conditions

PCR conditions were optimised, in particular for concentrations of DMSO, MgCl₂, Primers and DNA quantities. For PCR using microorganism genomic DNA and soil DNA as template (50 to 200 ng), the PCR mix (50µl) contained 250µM of dNTP, 5 mM MgCl₂ final, 2,5% DMSO, 1X PCR buffer, 0,75 µM of each primer and 2,5 U of Taq DNA polymerase (Sigma) and sterile distilled water. For PCR using fosmid pooled DNA as template (100 to 500 ng), the PCR mix (50µl) contained 250µM of dNTP, 5 mM MgCl₂ final, 5% DMSO, 1X PCR buffer, 0,75 µM of each primer and 2,5 U of Taq DNA polymerase (Sigma) and sterile distilled water. For identification of positives clones in 96 microtiter plates, 25 µl of each bacteria culture were used as template and PCR conditions were the same as above. Thermocycling program was : a denaturation step at 96°C for 5 minutes; then 1 minute at 96°C, 65°C for 1 minute, 72°C for 1 minute. The first 7 cycles, the annealing temperature was lowered 1°C per cycle until 58°C was reached. A subsequent 40 cycles were carried out with the annealing temperature at 58°C. A final extension step was at 72°C for 7 minutes. For identification of positives clones in 96 microtiter plates, the first denaturation step of 96°C was during 8 minutes. The other steps were the same as described above. PCR reactions were performed with a PTC 200 thermocycler (MJ Research).

PCR products analysis

PCR products of about 465 bp were purified on agarose gel with gel extraction Kit (Qiagen) according to the manufacturer recommendations. First approach consisted in subcloning PCR products using the Topo PCR II kit (Invitrogen). Recombinant Plasmids were extracted using QIAprep plasmid extraction Kit (Qiagen) and sequenced with Forward and Reverse M13 primers with CEQ 2000 automated sequencer (Beckman Coulter). Second approach consisted in direct sequencing of PCR products. Sequencing data were compared with nucleic and proteic genbank database using BLAST program.

Sequencing of the identified fosmid insert DNA and sequence analysis

Fosmids inserts were sequenced using either a transposon-mediated and by shotgun subcloning approach. Transposition was realized by using (Transposition Kit) commercialized by Epicentre according to the manufacturer. For shotgun subcloning, transformants were grown for 16 hours at 37°C. Fosmid extraction was done by using
 5 the Nucleobond PC100 extraction kit (Macherey Nagel). DNA was partially restricted with *Sau3A* and sized on standard gel electrophoresis for fragments ranging from 1 to 3 Kbs and cloned into Bluescript vector according to Sambrook *et al.* (1989).

Sequence analysis was performed with the identification of ORFs by using Frameplot of the GC3. Each identified ORF was compared to gene databases by using BLAST
 10 program. PKS domains were determined by aligning obtained sequence versus already described PKS domains from domain databases.

Sequencing

Sequencing reactions were performed with 1 µg of DNA and 3.6 pmol of primer, using CEQ 2000 Dye Terminator Cycle Sequencing kit (Beckman Coulter) under conditions
 15 proposed by supplier. Ten µL of reaction products were precipitated using 4 µL of solution containing 1.5 M NaOAc, 50 mM EDTA and 60 µL cold 95% ethanol/dH₂O from -20 °C. The pellet was washed 2 times with 200 µL 70% ethanol/dH₂O, vacuum dried and dissolved in 40 µL sample loading solution (supplied in kit). Sequencing reactions were run on an CEQ 2000 sequencer (Beckman Coulter).

20

Plasmids construction and validation

Plasmid pP1 was constructed as follows. A 941 bp DNA fragment containing native promoter region and AA(3)IV gene was amplified by polymerase chain reaction (PCR) using primers AmF (d-CCCTAAGATCTGGTTCATGTGCAGCTCCATC, SEQ ID
 25 NO: 9) and AmR (d-TAGTACCCGGGGATCCAACGTCATCTCGTTCTCC, SEQ ID NO: 10). One hundred microlitre reaction were performed containing 0.1µM each of primers, 1 X Vent DNA polymerase buffer (NEB), 0.2 µM of each deoxyribonucleoside triphosphate (dNTP), 50 ng of the DNA template and 2U of Vent DNA polymerase

(NEB). PCR mixture was heated for 4 min at 94 °C in a PTC-200 thermocycler (Peltier) and cycled 25 x at 94 °C for 60 sec, 59 °C for 30 sec and at 72 °C for 70 sec. The final extension was performed at 72 °C for 7 min.

5 PCR product was purified using GFX DNA purification kit (Amersham), then digested by *Bgl* II and *Sma* I restriction enzymes. A 941 bp *Bgl*II/*Sma*I fragment was inserted into the *Bam* HI, *Sma* I sites of pMOD plasmid (EPICENTRE). DH10B *E. coli* was transformed with pP1 and subjected to apramicyne selection on LB agar plates. Six colonies surviving on apramicyne selection were grown in liquid LB media and final pPL1 candidates were thoroughly checked via PCR and restriction mapping.

10

To construct conjugative plasmid pPL1, 750 bp oriT DNA region from plasmid RP4 were amplified via PCR using primers oriTF (d-GCGGTAGATCTGTGATGTACTTCACCAGCTCC, SEQ ID NO: 11) and oriTR (TAGTACCCGGGGATCCGACGGATCTTTTCCGCTGCAT, SEQ ID NO: 12). PCR
15 conditions were as above. Amplified DNA was digested using *Bgl* II and *Sma*I restriction enzymes. *Bgl*II/*Sma*I DNA fragment was subjected to purification after gel electrophoresis on 0.7% agarose. Purified fragment was ligated into pP1 plasmid digested by *Bam* HI and *Sma* I restriction enzymes.

20 ϕ C31 integrase gene and attachment site (attP) was amplified via PCR using primers Fint (d-AACAAAGATCTCCCGTACTGACGGACACACCG, SEQ ID NO: 13) and PJ (d-CGGGTGTCTCGCATCGCCGCT, SEQ ID NO: 14). Amplified DNA fragment was purified by GFX kit (Amersham) and phosphorylated using T4 polynucleotide kinase (NEB) under conditions recommended by the enzyme manufacturer. Phosphorilated
25 DNA fragment was cloned in to pPL1 vector opened with *Sma*I restriction enzyme (NEB) and dephosphorilated by calf alkaline phosphatase (NEB). DH10B *E. coli* was transformed with ligation mixture using Bio Rad Pulsing apparatus and protocols provided by Bio-Rad. Twelve transformants were analyzed by PCR for the presence of

integrase gene. Orientation of integrase gene was verified by restriction analysis using *Bgl* II and *Eco*RI restriction enzymes. Resulting plasmid was named pPAOI6.

To construct pPAOI6-A plasmid, pPAOI6 plasmid was digested with *Eco*RI and *Bgl*III restriction enzymes followed by digestion by Bean mung nuclease (NEB). Linearised
5 plasmid was self ligated and transformed into DH10B cells.

Plasmid preparation of fosmid DNA

Fosmid and BAC DNA for sequencing was prepared by using the Nucleobond AX kit (Macherey-Nagel), following protocol for BACs, Cosmid as specified by manufacturer.

Mutagenesis

10 Transposon Tn-pPAOI6 was prepared by digestion of pPAOI6 plasmid using *Pvu*II restriction enzymes, followed separation on agarose gel and purification of fragment containing transposon from gel using Qiagen kit. The same molar ratio of transposon and corresponding fosmid was used for mutagenesis in vitro using Tn5 transposase (Epicentre) and conditions specified by manufacturer. We transformed aliquot of the
15 transposed mixture by electroporation into competent DH10B *E. coli* strain.

Conjugation *E. coli*- *Streptomyces lividans* TK24

Conjugation experiments were done using 6×10^6 *E. coli* S17.1 cells containing conjugative plasmids or fosmids. The *E. coli* cells were grown in LB media with adequate antibiotic. The cells were collected by centrifugation, washed two times using same
20 volume of LB media and concentrated to 10^8 cells/ml and overlaid on LB plates containing 2×10^6 pregeminated *Streptomyces lividans* TK24 spores. The cells mixture were grown overnight at 30°C and *E. coli* cells were washed three times using 2 ml of LB media. The plates were overlaid using top agar containing NAL (nalidixic acid) and the appropriate antibiotic. Plates were incubated for 4 days at 30°C and transformant
25 streptomyces colonies were isolated on HT medium (Pridham *et al.* 1957) containing NAL and the same appropriate antibiotic.

Results

Construction of a Transposon Tn <Apra>.

E. coli aminoglycoside -(3)- acetyl transferase IV gene (aa(3)IV) was amplified by PCR and cloned in pMOD vector (Epicentre) . Advantage of this selective marker allow
5 positive selection in *E. coli* and in *Streptomyces lividans*. Transposon can be used for insertional inactivation in vitro using purified transposase Tn5 (Epicentre). The structure of the pP1 constructed vector and transposon was shown on Figure 1a,1b.

Construction of a conjugative plasmid-transposon.

Conjugative vector, transposon was constructed by cloning origin of transfer from
10 plasmid RP4 into pP1 vector producing pPL1 vector (Figure 2a, 2b). The origin of transfer was cloned in such orientation that the selective aa(3)IV gene is the last transferred during conjugation. PPL1 vector was introduced in to specific *E. coli* S17.1 strain that carry RP4 plasmid integrated in to chromosome. In conjugation experiment
15 between donor strain S17.1 carrying pPL1 plasmid and DH10 *E. coli* receptor we obtained DH10B strain carrying pPL1 plasmid. This data shows that cloned oriT fragment is functional in the pPL1 plasmid. Plasmid pPL1 can be used for DNA cloning, gene inactivation by homologous recombination. Cloned genes or part of the gene cloned could be then transferred by conjugation in to another host. Another advantage of
20 this vector is conjugative transposon that can be excised from vector and inserted randomly in vivo in to another DNA molecule by purified Tn 5 transposase.

Construction of a conjugative, site specific integrative plasmid- transposon for horizontal gene transfer between *E. coli* and *Streptomyces* strains.

PPL1 plasmid was used to clone an integrase gene from phage ϕ C31, resulting a plasmid pPAOI6 (Figure 3a, 3b). We tested several clones for horizontal transfer between *E. coli*
25 S17.1 strain and *Streptomyces lividans* TK24 strain. The best transfer was obtained for plasmid pPAOI6 were orientation of integrase gene is in opposite orientation to the gene for resistance to apramycine. Conjugative transfer of pPAOI6 gene in to *S. lividans*

strain is confirmed by resistance to apramycin or G418. Additional confirmation of transfer was obtained using PCR method. We were able to amplify 2 kb insert using specific primers for the ϕ C31 integrase gene, and no PCR amplification was obtained for control *S. lividans* TK24 strain.

- 5 The goal of this construction was to produce transposon that is further used in functional analysis of the metagenomic DNA library from the soil that was constructed in a laboratory (Figures 4a, 4b, 4c).

Functional analysis of the metagenomic DNA library from soil.

The fosmids library consists of 120 512 clones, containing ~ 40 kb inserts of soil DNA.

- 10 The library contains approximately 4.8 Gbps of the DNA cloned from soil. Ten percents of the library was screened by using a PCR approach for the presence of the genes involved in production of secondary metabolites (PKS). Using gene-module specific set of primers we were able to identified positive clones organized in microtiter plates (96 wells). Sequences (based on PCR products) obtained from fifteen randomly positive
15 clones indicate that the DNA library contains very little sequence redundancy limited to one and that the sequences were found to be new and very diverse in comparison to gene databases (data not shown).

Two fosmids DNA was prepared from two positive clones (FS3-124 and FS3-135) and analyzed by sequencing. DNA analysis *in silico* shows high G+C contents of 72% and

- 20 69% respectively of the cloned inserts and presence of cluster genes that could be involved in biosynthesis of secondary metabolites (Fig 5 a and b). No specific phenotype was observed for the two clones in *E. coli*. We employed pPAOI6 transposon mutagenesis to produce conjugative mutants. Transposon mutants FS3-124::pPAOI6 and FS3-135::pPAOI6 were isolated using apramycin as selective antibiotic. Obtained
25 mutants were then tested on LB plates containing chloramphenicol. About 1% of the tested clones are chloramphenicol sensitive. These clones contain transposon inserted into locus encoding chloramphenicol resistance gene and not into cloned DNA insert. ApraR and ChloS transposon mutants are then used for horizontal gene transfer into

Streptomyces lividans TK24 strain. Fosmid DNA was prepared from mutants and transformed into *E. coli* S17.1 strain. Horizontal gene transfer between *E. coli* S17.1 and *Streptomyces lividans* was done due to inserted pPAOI6 transposon. Transconjugants of the *Streptomyces lividans* were tested by PCR to confirm gene transfer and integration of the conjugative fosmid in to *S. lividans* chromosome. Both transconjugants showed an increase in doubling time, morphological modifications and pigments production in comparison to the control (Figure 6).

References

- 5 Brau, B., Pliz, U., and Pipersberg, W. 1984. Genes for gentamicin-(3)-N-acetyltransferases III and IV: I. Nucleotide sequence of the AAC(3)-IV gene and possible involvement of an IS 140 element in its expression. *Mol Gen Genet.* 193 (1), 179-187.
- Guiney, D.G., Yakobson, E., (1983). Location and nucleotide sequence of the transfer origin of the broad host range plasmid RK2. *Proc Natl Acad Sci USA*, 80: 3595-8.
- 10 T. Kaeberlein, K. Lewis, Epstein S. S. 2002. Isolating "Uncultivable" Microorganisms in Pure Culture in a Simulated Natural Environment. *Science* 296, 1127-1129.
- Mazodier, P., Thompson, C., 1989. Intergeneric conjugation between *E. coli* and
15 *Streptomyces* species. *J. Bacteriol.* 171, 3583-3585.
- Pace, N.R. (1997). A molecular view of microbial diversity and the biosphere. *Science* 276: 734-740.
- Pridham, T.G., Anderson P., Foley, C., Lindenfelser, L.A., Hesseltine, C.W., and
20 Benetdict R.C. (1957). A selection of media for maintenance and taxonomic study of *streptomyces*. *Antibiotics Annual* 1956-1957, 947-953.
- Voeykova T, Emelyanova L, Tabakov V, Mkrtumyan N. 1998. Transfer of plasmid pTO1 from *Escherichia coli* to various representatives of the order *Actinomycetales* by
25 intergeneric conjugation. *FEMS Microbiol Lett.* 162(1):47-52.
- Sambrook, J., Fritsch E.F. and Maniatis T. 1989. *Molecular cloning: a laboratory manual*, 2nd edition. Cold Spring Harbor Laboratory, Cold Spring Harbor NY.

Torsvik, V., Goksoyr, J. Daae, F. L. 1990. High diversity in DNA of soil bacteria. *Appl. Environ Microbiol.* 56, 782-787.

5 Torsvik, V., Sorheim, R. and Goksoyr, J. Total bacterial diversity in soil and sediment communities a review. *J. Industr. Microbiol.* 17, 170-178.

Torsvik, V., Daae, F.L, Sandaa, R.A., Overas, L. 1998. Novel techniques for analysing microbial diversity in natural and perturbed environments. *J. Biotechnol.* 64. 53-62.

10

Zaehner, H. and Fiedler, F. P. 1995. Fifty years of antimicrobials: past perspectives and future trends. In *The need for New antibiotics: Possible Ways Forward. Fifty -Third Symposium of the Society for General Microbiology*, Bath. England, UK, (Hunter, P.A., Darby, G.K. & Russell, N.J., eds), pp. 67-84.

15

Zechner, E.L., de la Cruz, F., Eisenbrandt, R., Grahn, A.M., Koraimann, G., Lanka, E., Muth, G., Pansegrau, W., Thomas, C.M., Wilkins, B.M. and Zatyka, M. (2000) Conjugative-DNA transfer process. In Thomas, C.M. (ed.) *The horizontal gene pool.* Harwood Academic Publishers.

20

CLAIMS

1. A method of analysing a library of polynucleotides, said polynucleotides being contained in cloning vectors having a particular host range, the method comprising (i) selecting cloning vectors in the library which contain a polynucleotide having a particular characteristic, (ii) modifying said selected cloning vectors to allow a transfer of said vectors and/or expression of the polynucleotide which they contain into a selected host cell, and (iii) analysing the polynucleotides contained in said modified vectors upon transfer of said modified vectors into said selected host cell.
2. The method of claim 1, wherein the library comprises a plurality of unknown polynucleotides.
3. The method of claim 1 or 2, wherein the library comprises a plurality of environmental DNA fragments.
4. The method of any one of the preceding claims, wherein the cloning vectors of the library are *E. coli* cloning vectors, preferably cosmids, P1 or BAC.
5. The method of any one of the preceding claims, wherein the selected vectors are modified by targeted insertion, into the vector, of a target polynucleotide construct.
6. The method of claim 5, wherein the targeted insertion is performed in a region of the vector distinct from the polynucleotide.
7. The method of claim 5 or 6, wherein the target polynucleotide construct comprises an origin of transfer functional in the selected host cell.

8. The method of claim 7, wherein the origin of transfer is functional in *Streptomyces* host cells.

9. The method of claim 8, wherein the origin of transfer is selected from RP4, pTiC58, F, RSF1010, ColE1 and R6K(α).

10. The method of any one of claims 5 to 9, wherein the target polynucleotide construct comprises an integrase functional in the selected host cell.

11. The method of claim 10, wherein the integrase is selected from ϕ C31.

12. The method of any one of claims 5 to 11, wherein the target polynucleotide construct comprises a transcriptional promoter functional in the selected host cell.

13. The method of any one of claims 5 to 12, wherein the target polynucleotide construct is contained in a transposable nucleic acid construct.

14. The method of claim 13, wherein the transposable nucleic acid comprises, flanked by two inverted repeats, the target polynucleotide construct and a marker gene.

15. The method of any one of the preceding claims, wherein the cloning vector comprises a first marker gene and wherein, in step ii), the cloning vector is modified by:

. contacting in vitro, in the presence of a transposase, the selected cloning vectors with a transposon comprising, flanked by two inverted repeats, the target polynucleotide

construct and a second marker gene distinct from the first marker gene, and

. selecting the cloning vectors which have acquired the second marker gene and which have lost the first marker gene.

16. The method of any one of the preceding claims, wherein, in step (i), the cloning vectors are selected by molecular screening.

17. The method of any one of the preceding claims, wherein, in step (iii), the modified
5 cloning vectors are transferred into the selected host cell by conjugative transfer.

18. The method of any one of the preceding claims, wherein, in step (iii),
polynucleotides are analysed by determining the phenotype or properties of the host cell
upon transfer or expression of the modified vector.

10

19. A method for the identification or cloning of polynucleotides encoding a selected
phenotype, the method comprising (i) cloning environmental DNA fragments into *E.coli*
cloning vectors to produce a metagenomic library, (ii) identifying or selecting cloning
vectors in said library which contain DNA fragments having a particular characteristic of
15 interest, (iii) modifying the identified or selected cloning vectors into shuttle or
expression vectors for transfer or expression in a selected host cell, (iv) transferring or
expressing the modified cloning vectors into said selected host cell and (v) identifying or
cloning the DNA fragments contained in said modified cloning vectors which encode
said selected phenotype in said selected host cell.

20

20. A transposable nucleic acid construct, wherein said construct comprises an origin of
transfer flanked by two inverted repeats.

25

21. A library of polynucleotides, wherein said library comprises a plurality of
environmental DNA fragments cloned into cloning vectors, wherein said environmental
DNA fragments contain a common molecular characteristic and wherein said cloning
vectors are *E. coli* cloning vectors comprising a target polynucleotide construct allowing
transfer or expression of the environmental DNA into a selected host cell distinct from
E. coli.

22. A polynucleotide sequence comprising all or part of SEQ ID NOs: 1 or 2, or of their complementary strand.

5 23. An oligonucleotide comprising SEQ ID NO: 3 or 4.

ABSTRACT

The present invention relates to methods for the identification or cloning of polynucleotides encoding a selected phenotype, particularly from environmental DNA. In a specific embodiment, the method comprises (i) cloning environmental DNA fragments into *E.coli* cloning vectors to produce a metagenomic library, (ii) identifying or selecting cloning vectors in said library which contain DNA fragments having a particular characteristic of interest, (iii) modifying the identified or selected cloning vectors into shuttle or expression vectors for transfer or expression in a selected host cell, (iv) transferring or expressing the modified cloning vectors into said selected host cell and (v) identifying or cloning the DNA fragments contained in said modified cloning vectors which encode said selected phenotype in said selected host cell.

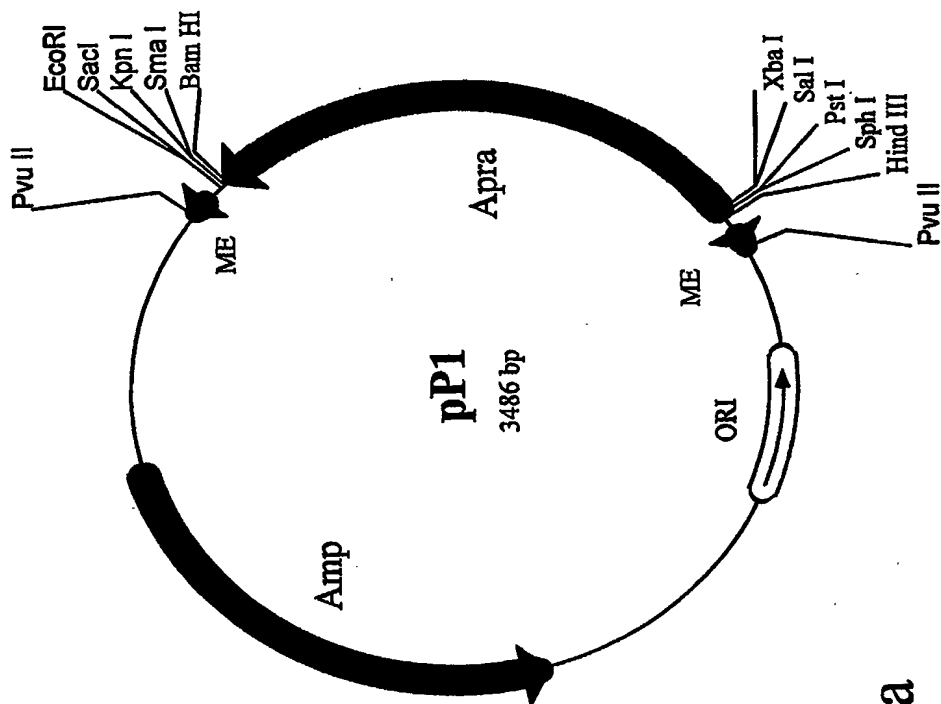


Figure 1a

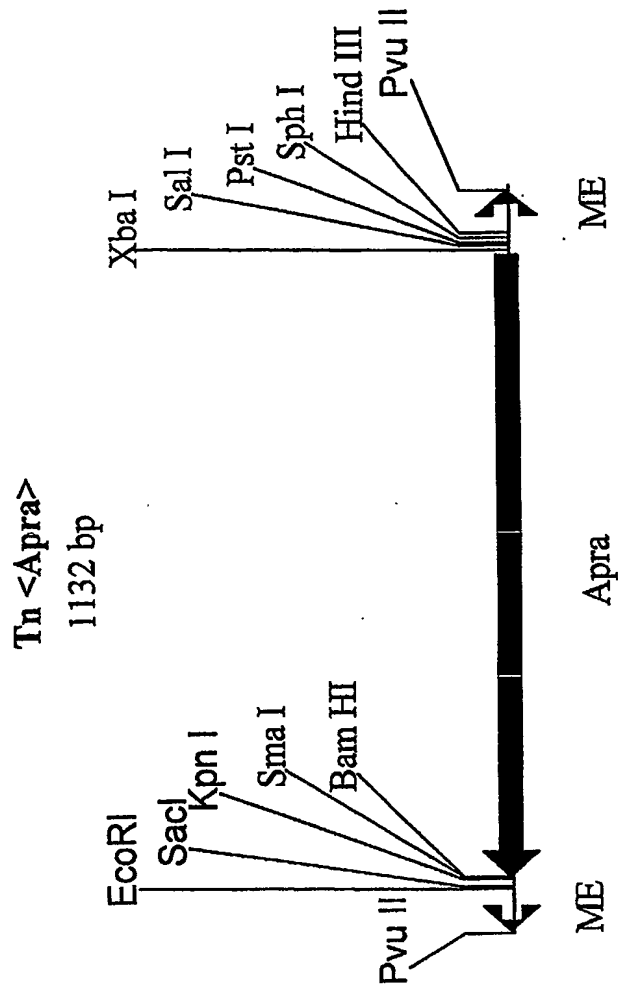


Figure 1b

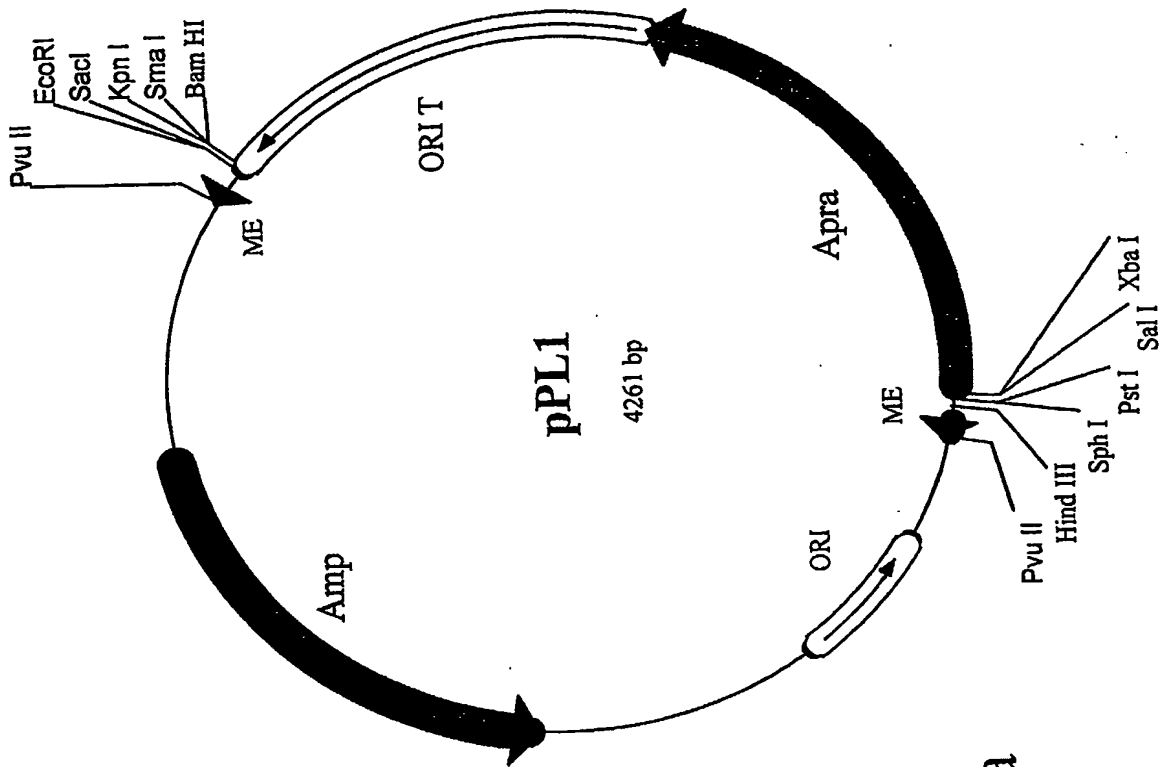


Figure 2a

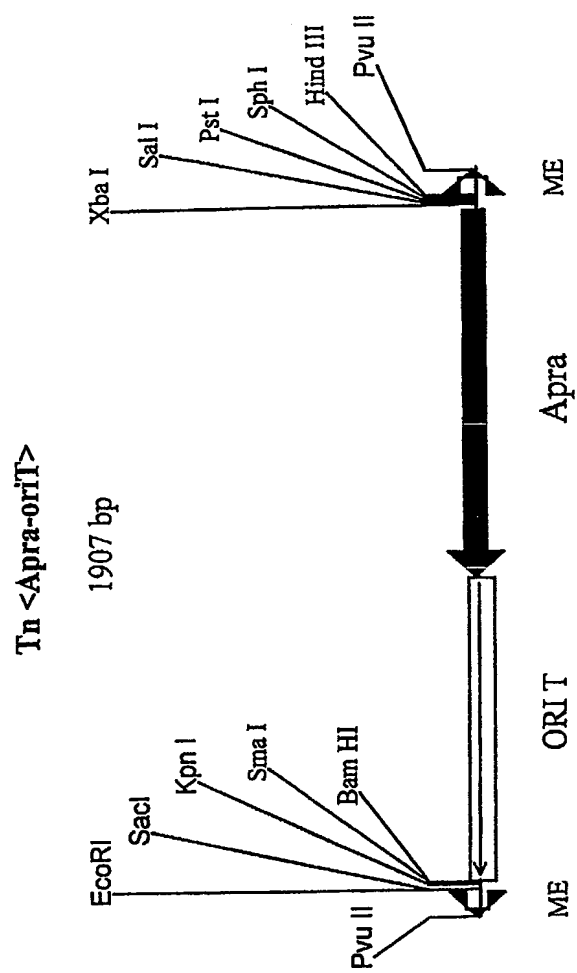


Figure 2b

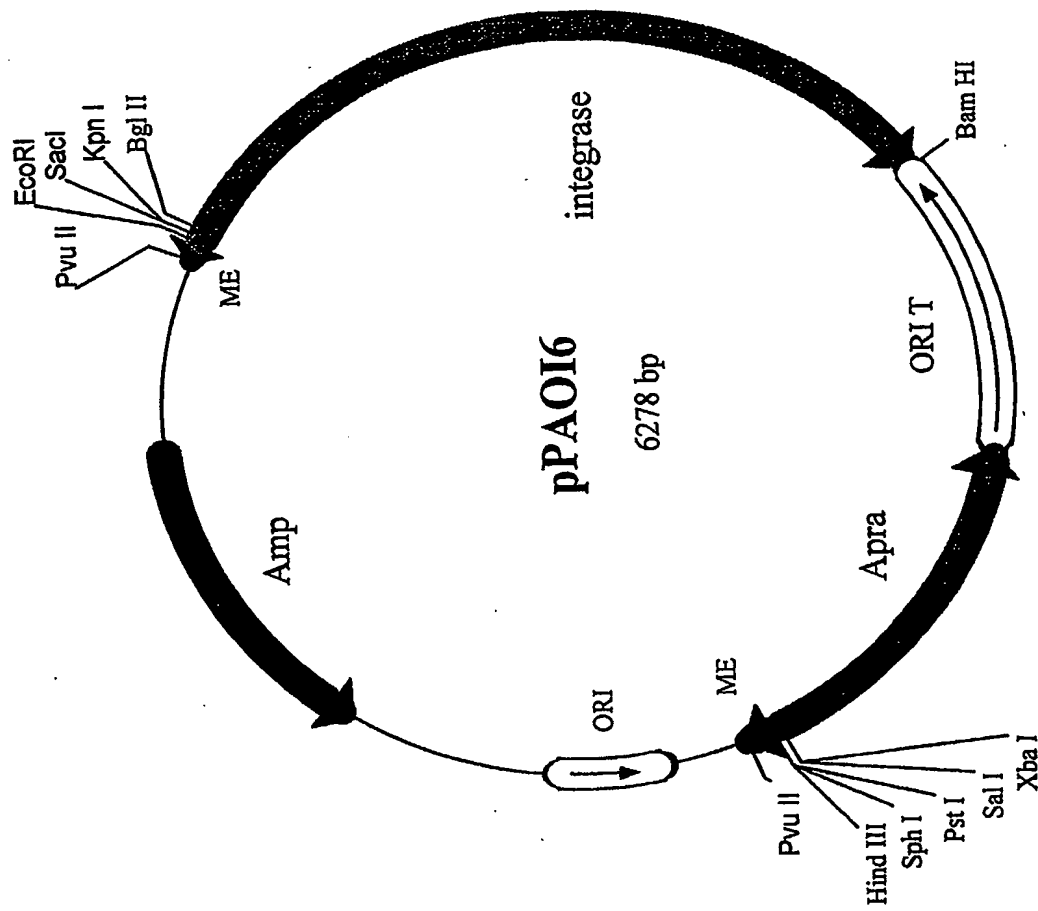


Figure 3a

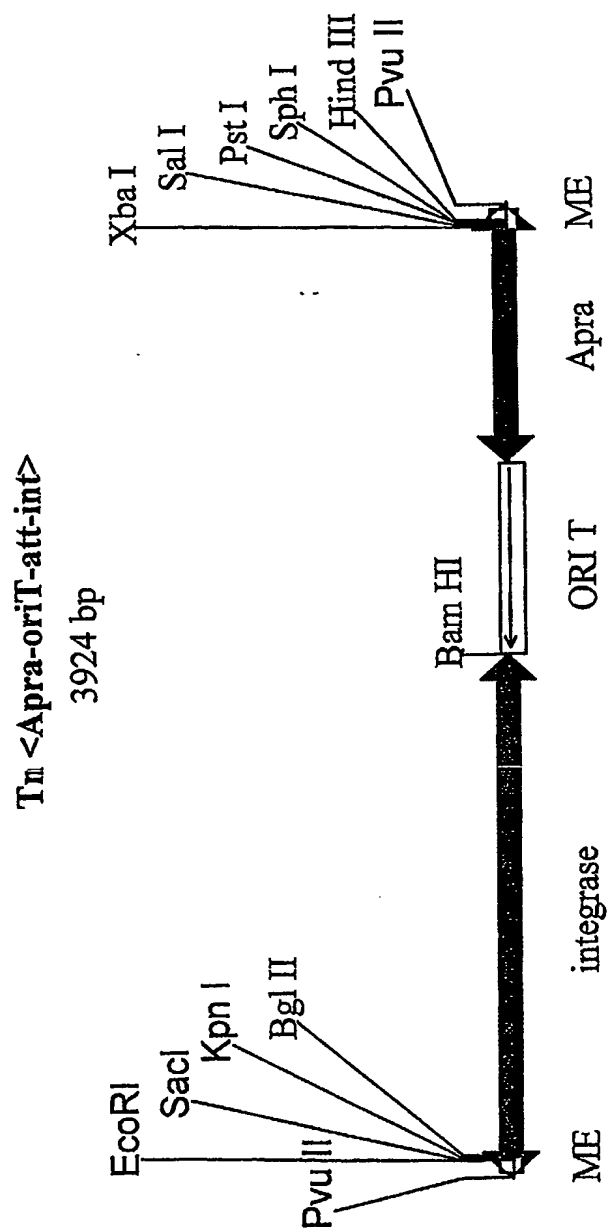


Figure 3b

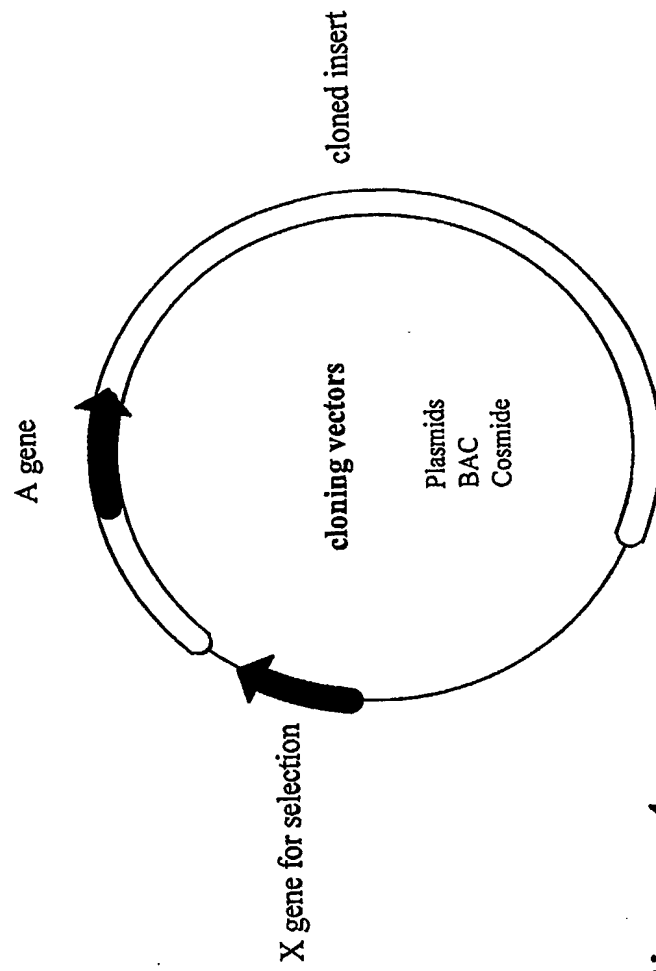


Figure 4a

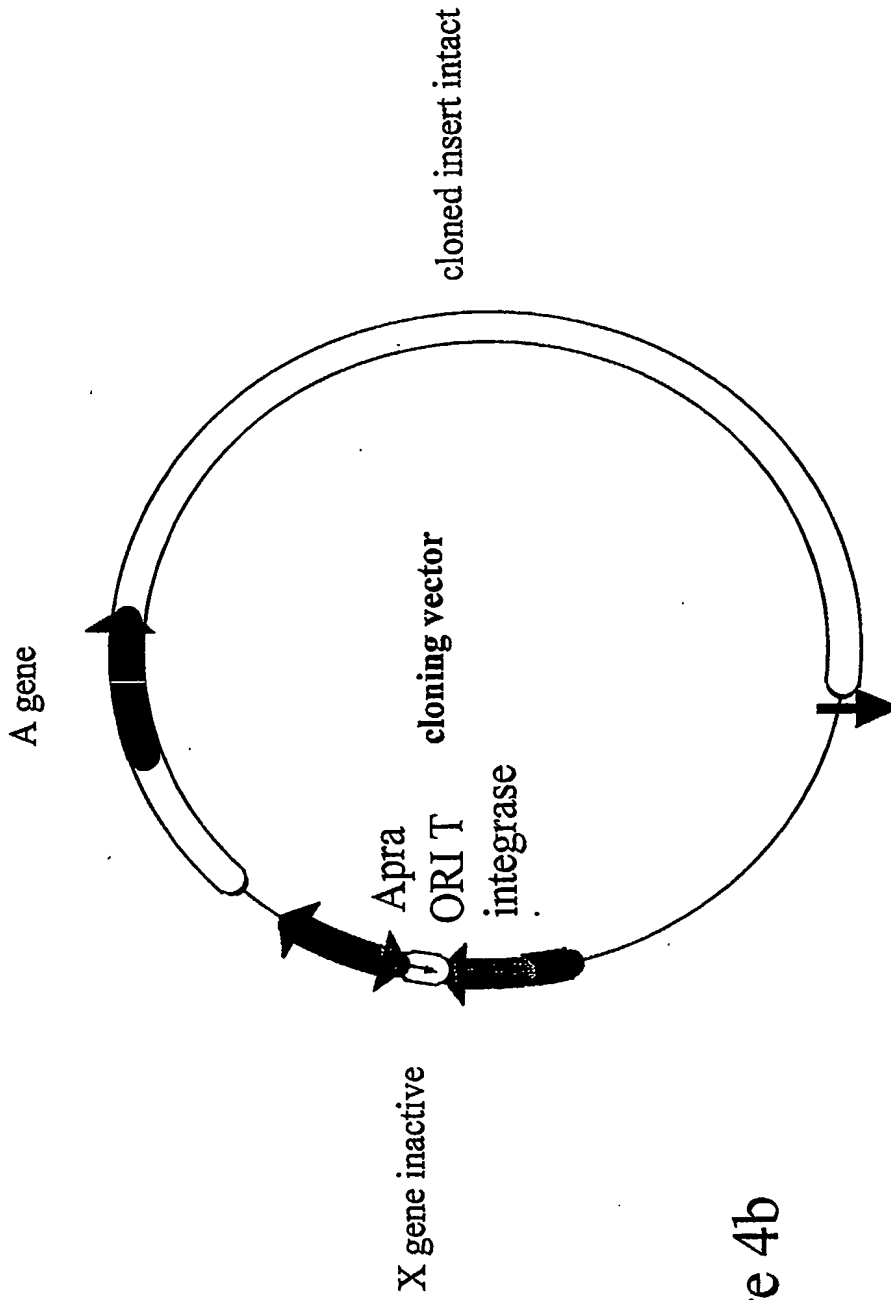


Figure 4b

Conjugative and integrative plasmid

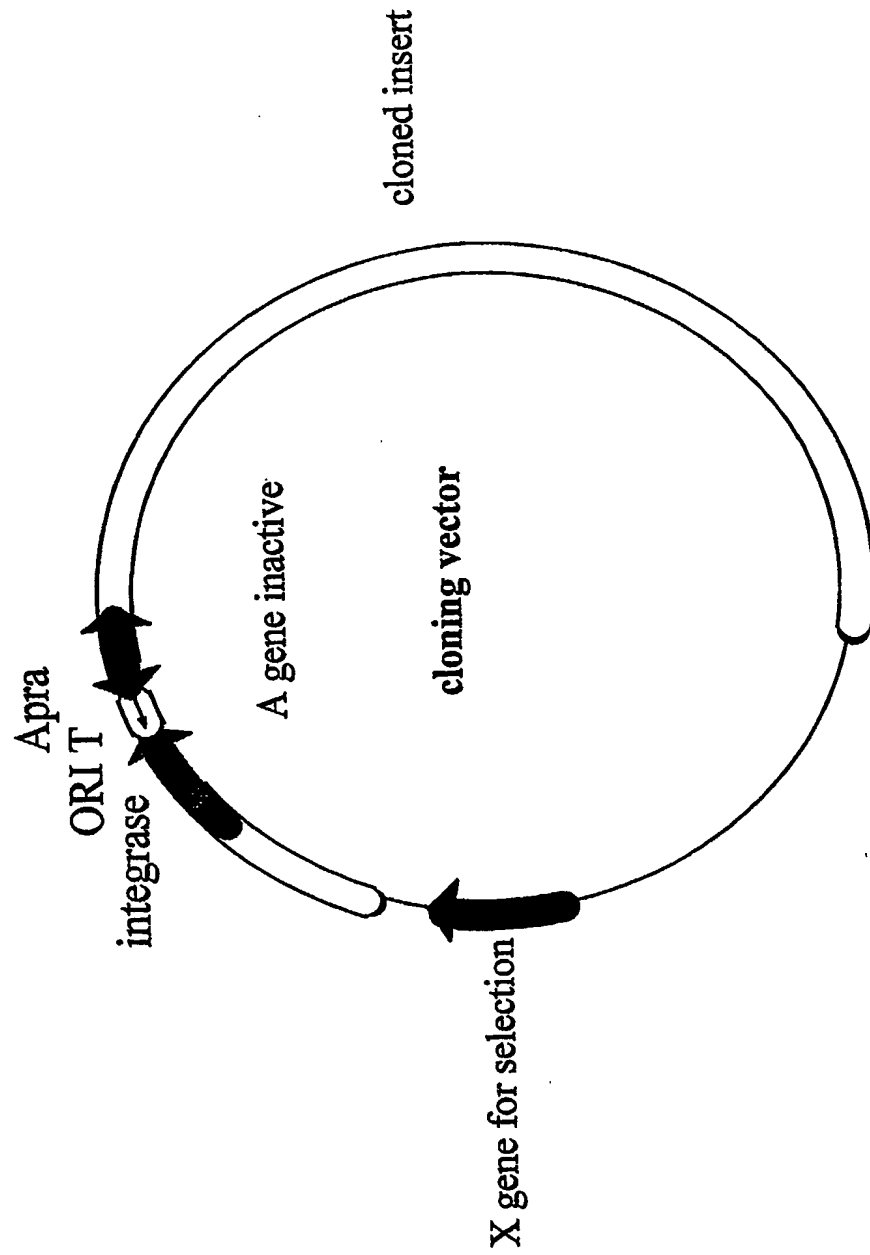


Figure 4c

```

CDS      76..1134
         /note="ABC_transportr"
         /gene="TAP2 PROTEIN"
         /blastp_match="Oryzias latipes"
         /blast_score= 0.002
CDS      complement(1096..2430)
         /note="none"
         /blastp_match="Anabaena sp"
         /gene="ALR1117 PROTEIN"
         /blast_score=2e-18
CDS      1178..1624
         /note="Gram_pos_anchor"
         /gene="CELL WALL SURFACE ANCHOR"
         /blast_score=1e-04
         /blastp_match="Streptococcus pneumoniae"
CDS      complement(2506..3567)
         /note="CONSERVED"
         /gene="HYPOTHETICAL PROTEIN"
         /blast_score=0.019
         /blastp_match="Deinococcus radiodurans"
CDS      complement(2906..4222)
         /note="glycosyl transferase"
         /gene="lipopolysaccharide"
         /blast_score=2e-23
CDS      complement(4092..5321)
         /note="glycosyl transferase"
         /gene="glycosyl transferase"
         /blast_score=1e-15
CDS      complement(6337..8502)
         /note="PUTATIVE"
         /gene="GLUTAMINE AMIDOTRANSFERASE"
         /blastp_match=" Bordetella bronchiseptica"
         /blast_score=1e-16
CDS      complement(8181..9530)
         /note="none"
         /gene="MEMBRANE PROTEIN"
         /blast_score=0.035
CDS      complement(9531..10721)
         /note="NOEC Transmembrane"
         /gene="NODULATION PROTEIN"
         /blastp_match="Azorhizobium caulinodans"
         /blast_score=3e-07
CDS      complement(10504..11274)
         /note="PUTATIVE"
         /gene="HYDROLASE"
         /blastp_match="Streptomyces coelicolor"
         /blast_score=4e-14
CDS      12874..13689
         /note="HYPOTHETICAL Meth-transf"
         /gene="PROTEIN PA1088"
         /blast_score=2e-06
         /blastp_match=" Pseudomonas aeruginosa"
CDS      14195..15976
         /note="PUTATIVE, Glyco_transf"
         /note=""
         /gene="LIPOPOLYSACCHARIDE BIOSYNTHESIS"
         /blast_score=2e-06
         /blastp_match="Vibrio cholerae"
CDS      15427..16512
         /note="PATHWAY: INNER CORE LIPOPOLYSACCHARIDE
         BIOSYNTHESIS"
         /gene="PHOSPHOHEPTOSE ISOMERASE "
         /blast_score=3e-17
         /blastp_match="Helicobacter pylori"
CDS      15579..16253

```

/note="none"
 /gene="PHOSPHOHEPTOSE ISOMERASE"
 /blast_score=2e-22
 /blastp_match="Neisseria meningitidis"
 complement(16505..17656)
 CDS
 /note="BIOSYNTHESIS PUTATIVE"
 /gene="LIPOPOLYSACCHARIDE"
 /blast_score=2e-17
 /blastp_match="Thermotoga maritima"
 /pfam_match="Glycos_transf_1"
 complement(17657..18697)
 CDS
 /note="none"
 /gene="ALR3073 PROTEIN"
 /blast_score=6e-27
 /blastp_match="Anabaena sp"
 complement(18615..19304)
 CDS
 /note="none"
 /gene="ALR4487 PROTEIN"
 /blast_score=8e-07
 /blastp_match="Anabaena sp"
 complement(19301..20596)
 CDS
 /note="ATP_GTP_A"
 /gene="ABC TRANSPORTER"
 /blast_score=3e-61
 /blastp_match="Synechocystis sp"
 complement(20535..21476)
 CDS
 /note="PERMEASE COMPONENT"
 /gene="POLYSACCHARIDE ABC TRANSPORTER"
 /blast_score=6e-41
 /blastp_match="Clostridium acetobutylicum"
 complement(22025..22951)
 CDS
 /note="involved in the synthesis of a polysaccharide capsule ?"
 /gene="32.3 KDA PROTEIN"
 /blast_score=3e-17
 /blastp_match="Sphingomonas sp"
 complement(23155..26523)
 CDS
 /note="peptide syntase"
 /gene="mcyA, mcyB and mcyC"
 /blastp_match="Microcystis aeruginosa"
 complement(26409..34433)
 CDS
 /note="polyketide syntase et peptide syntase"
 /gene="mcyD, mcyE, mcyF and mcyG"
 /blastp_match="Microcystis aeruginosa"
 complement(34418..37500)
 CDS
 /note="CYSTATIN"
 /gene="PEPTIDE SYNTHETASE"
 /blast_score=0.0
 /blastp_match="Anabaena sp"
 complement(35359..37500)
 CDS
 /note="gene cluster"
 /gene="nostopeptolide biosynthetic"
 /blast_score=2e-41
 /blastp_match="Nostoc sp"

Sequence 37500 BP; 6199 A; 12698 C; 12769 G; 5834 T; 0 other;

gacggtg	ggcctcttcg	ctattacgcc	agctggcgaa	agggggatgt	gctgcaaggc	60
gattaagt	ggtaacgcc	gggttttccc	agtcacgacg	ttgtaaaacg	acggccagt	120
aattgtaata	cgactcacta	tagggcgcaat	tgcagctcgg	taccggggga	tcccacgtac	180
cacgagctca	tctggaagag	cccgtccag	acacccgagg	aatggaagcg	cggcgtcgtg	240
agaccgtaca	cgcagaagcg	tctcgtcgcc	ttcggtgact	ggcttgcgac	gacgatgaag	300
ctggatctca	ccaggatggt	cgccgcaggc	agctcgatgg	gcggatcggg	cgcgatcatg	360
ctcgcgattc	gctatcccgc	acgattcgcg	tggaccgtgt	cgtgggtcgg	cgctccacgtg	420
cccgccgact	ctccgcaatt	cacatcgctg	tacgagctgg	tgtacggccg	gcccgacttg	480
aaggtgccgt	tcgagaacgg	cacgcccgtc	tgggacatt	ttagtgcagt	ctgggtacctg	540
gggcagcatc	cggagcagga	catcggttc	atcacgttct	cgaacggcaa	gaacgactcg	600
gcgacggct	ggcgccaggc	cgctgaattc	ctgaagacgc	tgcaggagac	caggcaaccg	660
cacctgttcg	tctggggcca	ggaggacac	ggccagcgcg	cgaagatgcc	ggaaggtggc	720

ggcgaacggg	agatgccgct	cgacctcaga	acgaaccaga	gcctgccggc	gttcagccgt
tgcacgctcg	acgacgatcc	gggagacggg	tcggattcga	gcggcgctcc	ggccgggag
atcaactcct	acgtcacctg	gcagcctgac	tcgatcgctg	acgaaccggg	ccgatggagc
gtcgatcatca	agctggcgaa	tcgaccgccc	cggtcgcccc	ccgcagtcga	cgtcacgccc
cgggcggtgc	agcagttcaa	gctcaagccc	ggcgatcagg	tgacatggac	gaactccgcg
ggcaacaccg	tcgtgcagcg	cgggcgaagcg	gtcgcggaac	ggtgggggtct	cgtcacgctc
ccgcagatgc	aggtctcaac	gggtggaaac	aggatcctgg	tgagcaggaa	ctaggcgggc
ggcgctcctcg	acgagccggc	gcatcgggcg	cagttgatgt	gcccacgaaa	gctggctgag
gtccgcgcg	gcggcgatgt	cccgtcgcg	ctcgcgggcg	tttgccagca	aatccagcac
agccgcccgg	aacgcctcga	ctgattccgc	cacggagcac	gcccggcggtg	cgatgggtgg
aagtcacctc	gcgacggccg	gggtcacgac	gttcgggagc	cccgcggcaa	tcgcttcgag
cactttgttc	tgacgcctc	tcgcatcg	aatcgggcg	accgcgatcg	ccgatcgcca
gagatactcg	cgtacctcgg	ggaccgccc	ggtcacttca	atcgtgggat	cgcccgcgag
tgcccgacc	gcggcggtcg	ggtccatgcc	cacgagcatc	agccgcgctg	ccgggcccgg
gcgaaccacg	gacggccaga	cctctcgcg	gaaccagagc	gcccgcgctg	cggtgggctc
gtagccgaac	acgcgcgaga	agacgacccg	cgcttcggcg	gcccgggtg	ctttcgggcg
aaacatctcg	acgtcgatgc	cggttcggcg	gaccgagccc	gtgaggccgc	ccgtgacttc
ttcgagcagc	aggcgctcgc	gctcgctcac	gaccgtgggtc	gcccggcgcg	gcccgcacgc
gactttctcg	aagcgggagc	agcggccggc	ctcgcgggcg	aacagccagc	ctttcgggca
cgccgcgctc	gacgccagcg	cgcgccatct	ctcgaggtcg	acgtccacca	tgtccaggac
gaaggggatc	ccgcccaggga	tcgggtcgca	ggcgtagcgg	gcatcccgcc	tgcatagcgc
gagcaaccg	tcggggcgcg	tcggcgcaat	ccgtcggttc	agcacctggt	agatctccgg
cgaatgcagc	aggacgtcg	tgaggggtcg	atcgccccgg	agcgccagac	cgctcgcgat
gagattccgg	ccgcggcgga	cgcgccagac	gtcggtcgac	gcccgtcacgc	cgggcgagatc
cgctcgatgc	gaccactcct	cgctcgctcg	cgcgagcgag	acgaggtgca	cgtcggcgga
tcgtcgagcg	gtgtgaatga	gatgaaacgc	gcgaatgcga	tcgccccggt	tcggggcgta
gggcagccgg	tgtgtgagca	gaaggactct	caccgtcgcg	catccgatcg	ctctcggtatg
acgcggggcg	gtacgcccgc	ggcgatcacg	cgggcgggaa	tcggcgcggt	cacgaccgac
ccgcgcgcaa	tcacgggaatc	cggaccgacg	tcagccatga	ccacggcggg	gctgccgatc
caggtgccgg	cgccgatccg	caccatccgc	ggtgtgcctg	gctgctcacg	gatcgggcgc
gtgaggtcgc	tcgtgcccgtg	cggtgtccgg	ccgctgggga	tgtaggacgc	ggcgccgacc
aggacatcgc	gctccagggtg	gaccagcccg	agatggcatc	cggttcccac	gtacacatgc
tcgtcgagct	tcgcgcgggt	ctgcgaaaag	atggtgccga	aggcgaccgt	gaccgacggg
tcgcagtggt	caagcacgcg	gcccaggaaa	gcccgcgcga	ggtattgtcc	gatgacggcg
gggatgagcg	agagccactg	cgctcgaccct	tcgagggccc	ggtcgggcgcc	gaggaagggc
cgccgcacgg	agtaggagag	caaggcgggc	aggacgagca	ccagcgcgat	gcccagggcc
agtcctctgc	cggcgtcctt	gaggtcacg	cccactgcag	cgcgcgcccc	gcgtcggaac
gggtcccgta	tcgtgtgacg	agcgagtcgt	agagcgcttc	cactttgcgc	atccgcggct
cgaacgataa	ctcttcttctg	atccggcgct	gcgcgcgaat	cgacgggtg	cgtcggggcg
cgggattggc	catcgccctg	tcgatcgcg	ccgcgaggcc	tgacgcccgg	ccgggttgaa
tgaccagggc	gtgcgtgccc	ttcgacacga	gctcgcccg	gccccggcc	gcgggtcgga
cgatgggtgt	ctcgaaacgc	atggcttcga	gcacggcggt	ggcgctccct	tcgtagtcgg
acgactgcac	gagcaggtcc	agcccgatg	ggacggccga	gacgtcgga	gtgtggccga
ggaaacgcc	ggttccgggg	ccgagctcgc	ggcgggccctg	ttcctcgagc	ggacgcccga
ggctgcccgc	gcccggcgatg	acgagacgga	ggcggggcca	cttcgcctga	agggcggcac
aggttccgat	gagcaagtcg	aatcgtttct	gcggctcgag	acggccaacc	gcgcggatga
cgaagtgcgt	ggcctcgaat	cccagtcctt	cacgcgcgcg	ggcctcagat	cggtggatcg
gtcggaagc	atggtgggtg	atgccattcg	ggatcgtagc	gaccgatcc	ggccgtccgc
ctttggccag	gagctcccgc	cgaatctgtt	cggacacggc	aatgacggca	gggaagcggg
cgagaatccg	gcgatcgagc	gggtagtaga	cggtgcttgc	gcgaacgaa	tggccggctc
atccatgggc	cggtggccagc	ggaatcacgc	gctcggcgcg	tcccaggagc	agggcgagga
gatcggtcct	gtactcgtgc	gcgtgcacga	tgctcgatgc	cagatctttc	acgagcgcg
ggagcttcga	ccagattccg	gtgtcgaacg	aatgcttctc	ccggacctcg	acgtagtcga
tgccggccgc	tttcgcccgt	tcgtccatgc	cgaacaccgg	atcgcgctgg	tcgcggtatg
agcagacggg	gatcgcgtag	ttggagcgat	cggcctgcgc	cgttccgagc	agaatcgctc
ttccgggtcc	gcccggcggtg	ccccgcacgc	tgcgaggtc	gagcacgcgc	acgggacgcg
cgattcccg	cctaggcgcg	gcgtctggca	gcatggcccc	ccgggcccgc	gagcgcgatg
cccggggcga	tatcgtagct	gcccggccacg	acgagcgact	gtgcgatctc	gacgagtcgc
actgcgcgcg	gcccggcgccg	gatcatcgcg	gctgcgcgca	agccggcgaa	accggcaagc
agcaggagcg	cggtaccgac	aatccaccac	ccgcgcgctg	gccacagcgc	gaggcccagc
ggaatcgaga	cgagcgccag	cagatcgagg	accgggatga	tcacgctggg	cagatccgcg
atggcgagcc	ggtgacggaa	gctcaegcgc	aggttgtcgc	ggccgcgcca	cagctcgccg
aaaaaacacg	cgcgaaagcga	tcggggatcg	ccctggtgca	cgctgcggag	ccgcgggtcc
gacaggatgc	ggccgcgcgc	ccggcgcgagc	cgctggcaga	tatccacgtc	ttcacaggtc
tcgaggctgg	tatcgaagcc	accggcgcg	tcgaaggcct	cgcgcttcac	ggcaagattg
ccgctcgggg	gccagtcac	gtcttcgggtg	ctctgcccac	gacggcgaaa	cgagtcgtac
gcgcgctgca	cccagttgcc	gtcaccggga	gcttcgtact	gcgcgcggcg	agcgacgacc
cccgggtct	ggagcagatc	gacggcgacc	accggccagc	gcgggtcgat	ctcatgatcg
gcgtcaacga	acgccagcac	gtcgccctcc	gcccagccgg	ccgcgcgggt	gcgaagcgct

780 12/35

840
900
960
1020
1080
1140
1200
1260
1320
1380
1440
1500
1560
1620
1680
1740
1800
1860
1920
1980
2040
2100
2160
2220
2280
2340
2400
2460
2520
2580
2640
2700
2760
2820
2880
2940
3000
3060
3120
3180
3240
3300
3360
3420
3480
3540
3600
3660
3720
3780
3840
3900
3960
4020
4080
4140
4200
4260
4320
4380
4440
4500
4560
4620
4680
4740
4800
4860

gccacgttca cgtcgggcag cgacagcact cgcgcggatc cgcgctctgc caccggcagc
 gagtcgtcag tccagccggt gtcgcgcacg accagctctgt accagccage cggcgccctcg
 gtgcgatgaa tgcagccagc gcaacggctg aggtgcgcgc cgcgcttctt gacggggatg
 acgaacgaca cgcgcggcgc caccggcaatt ttcgcggtca tccgattctgt agcgcggccg
 gctttcgtag cgcggcccg gttggggggc gcgagaactc gaacgcgcgc agatccggcg
 cccctgctct cggccgcgca tccgaatcga cggcgccctg cggaaatctcg atgccggcgt
 cgatcgccct cctggcggct gccgacaggt gcagatcgca cgaggcccg tttgcgaacc
 actctgcggc ggcgctcgtt acgttgctcg tccgctccgc ctaccgcgc ttaccggcgca
 ggaatcgccg gctcgtcagg ttgttgcgga tgagcccggt cgtcacggga aaccggacgc
 cgatcgccca cggcgctcgc gtgcggagcg atcgcttctg cggccattc gttgtgctcg atccggaagt
 cgggtggacgc gttggcttcg atcgcttctg cggccattc gttcaggtt gctcagctcagt
 tgttcgcgac gatgcgcgc ttgtgatcga tctgtctctc gccgtcgct gcgtagggcg
 tgatccccgc ggccaggccg atggcgatgc ccctgaacga atcgacgat aggttccgtt
 cgatgagcgt gtccgtcgag ttccgccagg cgaggatcgc cggcccggt cgcagccgc
 ccgattgcgg gccgcgaatc cgcagggaaga cgttgtcgcg gatgaccac cccggggccg
 ccaggatgtc caccgcgtcg gtgtaatcgg accggcgct cgtcgtgtac cgaagcgcg
 agcacgccac gaggccatcg tccgcgaagc gtcgctcagt gccagcgct cctttgagca
 attgctgtcc cgcactgatg agctggacgt tgtgacccga cgtggccgac cgtcagatcc
 cgccgcgcac ctgaatcgcg ttgtaccgga cgtggccgac gccggtcatg ccgcccgcg
 cgtcagtgcc gccaacggag agcgcacgc cgatcgctc ggcctccggag caccgtgccc
 gaatgacgac gcggtcccga tccgcggagc ggtctccggag cgttctcgag cagaatcgt
 tgctggccat ccggtcagac cgttactcgc agcagctcgg acgtccgcgc caccggtgacg
 tgcgcgcttc gagcgcgtg cgcgccatcg cagaagtgcg gcaccggcgc ggccgctg
 tccgcggcgc cgcgccatcg gactgcgcgc gccacgatga ggcgagtcag gccgccaccg
 gaagcgccac gaggccatcg aaccagagct cgaggtatgag caggatttga aggaggtagc
 gtgttcgcgc cgcgtgtcgt tccggggcga acagctcgcg ctccgacggt tccgggcgca
 tctcccgta cgcgtgtcgt tccggggcga acagctcgcg ctccgacggt tccgggcgca
 ccagcgatcg aagtgcgtg agtgacggta tctccgccag cggaccggcg tccggcgcg
 gttgacctg tccgaacact gcgcggggt gcccgccgc cagcaggacc gaagagatcg
 cgagtccgc acctgagca cgcacgtcca cgacggtcgc gctgagctcc gctgagctcc
 cgtgtcgtcg cgcctatccc agagatagct agtcggccgc gcccgcctcg gctgagctag
 gaacggcagc cgcacgtcca agagatagct agtcggccgc gcccgcctcg gctgagctag
 ccgatgatgc tccgaagat agtcggccgc gcccgcctcg gctgagctag
 gtcgaggagc gcgtgcagcg gtgtccagcg gcacatcgc cgtcatcgc gctctccgc
 tgtgaagagc gtgtccagcg gcacatcgc cgtcatcgc gctctccgc
 gtcgatgaac tccggcgccg cgtctccgc ccagcatctc cctgggttcg taccggtag
 cgccaggctc gtcttggcga cgtctccgc ccagcatctc cctgggttcg taccggtag
 gtcggcagg gtcggcgcca gcgtgtacgt ggccgcggga gctgaacgt agagatcggc
 cagcagccgc cggctgtcga gctgaacgt agagatcggc cgaagcccag aggtcctgag
 gagecactc tccagccgc gctgaacgt agagatcggc cgaagcccag aggtcctgag
 cgtcgtacgc ggtccgcggt agagatcggc cgaagcccag aggtcctgag
 gcctgacggc cagtcgtacg cagcagccag cgaagcccag aggtcctgag
 cagggtcttg tccgccagga ttgcgcgcg agtagatcgg atagatcgg
 gagagccgt tccggccgc cgcgcggat cgatcacggc cgtacaccaa agccaggaga
 cactcgctct cccagccgtt cgtacaccaa agccaggaga gttgtagaga tccgtcgagg
 catggcaggc cccagccgtt cgtacaccaa agccaggaga gttgtagaga tccgtcgagg
 cgaacaaac tccgtccgga cgcgcgctt cgcgcggccac gatctcgata gacgaccgtg
 cggccggctg acgtcgcgc gctgaacgt agagatcggc cgtgtgcacc cggctcgaaa
 gaccgcccat cgggtccgag gatctcgata gacgaccgtg cgtgtgcacc cggctcgaaa
 cgccgctgac gcggcgtcgg cgcggtcgaa ggcggtagat cccgcgggag cgcggcgcg
 gacgcccagg aggcggttca tccagacgacc cccgcgggag cgcggcgcg
 ggtgtcgtct gagcggccgg gtgcccagg gtgcccagg atctcgggtg tggacggcca
 gcaatcgaca agccaggaa gtgcccagg gtgcccagg atctcgggtg tggacggcca
 ccgcccagc agaattgcaat gagggcggtc tggacggcca ggcaggcgat ggcaggcgat
 gccaggggcg gatcccgctt cgagaccggc cgcgagggcc accgaaccaa gctcagccac
 aggaagtgat gcagggaacgc ccccttgccc caggcgccgt cccagctgtc gccggtcac
 cagcgctgtg gaaccgagc gtgaccgac cgttaaatcg gccggtcac gctgatcgcc
 agcgcgaccc tccaggaagt cagccggctc tgggcccagat cgtgatcgcc tcatgcgat
 tccgtcgcgt cgatgctcga gtcggacgag accgcgcgc cggccacgag cagcgctgag
 tccgtcatca cgagcgacat cgcgacgaag accgcgcgc cggccacgag cagcgctgag
 ttgcgcttc agtacaggaa atagacgag gcatagagc accccgatga gacgacgc
 cgggagtcag tgcgtatgcc cgtctgcac agcgtgacga tgagcggcac
 ttggccatt tctcgcctc cgtctgcac agcgtgacga tgagcggcac

atgccgacgg	cgacgccgtt	gttggtcaccg	agcgagtaga	cgctcgttgaa	gttcttgatg	9060
ccgggaatga	acaggagctg	accccagccc	tgcttggcgg	cctcgaaccc	gagggagagc	9120
gagatcacga	ggaagacgag	ccgcaaccgc	ttgacgtccg	tcgtgaggac	tgcgagcagg	9180
tacgtcatca	cggtggactt	catgaagtgc	atcatgtagc	cccacgcgaa	gtcctggtag	9240
ggcgacgcca	gcgtcgacga	caggctgtgg	agcccgaaga	aggcgaggag	cgccagccgc	9300
gcgtcgatcc	gcagcgccgtg	cccggacagg	aacgcgatcg	cgagcgtgta	catgcccgc	9360
aggagcgaga	tggtcatgtg	gatgagaaag	tcgctccaca	cccagagctc	cgcccggaag	9420
tacgcgatga	acaggtagaa	cagaatcgcg	tagaacggcc	cgcgacgcgc	gtggaaggcg	9480
ccgaaggcga	ggagggcgag	gacgaaagcg	gctcgaagca	ttgccgtgtg	ttacctgtta	9540
tcggtccccg	tggtccagag	cggtgtgcgc	cgggtctcct	ctacattgaa	cacctgttag	9600
agggccggaa	tcgaggtgaa	catgagcacc	acgaaaagaa	tcgccgacac	cacctgttag	9660
gcgaagaacc	cccgtgctt	gtagagcttc	tccggattct	gcacggggct	gttcggcagc	9720
atcccgatgt	gcaggtagta	ggcgaacatg	cccgcgcgca	ccggcgcgaa	caggatcagc	9780
tcgaggtgat	agcggacgat	gaagatgccg	ccgaagaggg	cgagcaggt	ggcgtagaac	9840
acctgctca	cgagcagccg	ctcctcggtg	tagaaggcaa	acgacttcgg	gtacgagcc	9900
gcgaccgacg	cggtgcccgt	gtgccggaac	tccgcgtacc	gcttcgtcgc	catgaagaac	9960
gcgcccacca	tccagtacga	gatcacgagc	gacatcgccg	gcacgcgatc	ctgaatcagc	10020
gggaaccagc	cgaggaggag	ccggacggcg	ttgttcgccc	actcgctcag	cacgtcgagg	10080
tacggccact	ccttcgtgcg	aatggggcgg	acgttgtagc	tgacgccgag	cacccagagc	10140
gccaggggcg	acagcgcgaa	gtatctgttc	acgagcagcg	cgagcacgaa	tccggccacg	10200
cccacgagga	tccactcggg	gtagcccgc	gcccgtctga	tcttcccgga	gggcaccggc	10260
cggtgcccgt	tctccggatg	cagcagatct	ctcgggccgt	ccaggagctc	gttcagcagc	10320
tagttgctcg	aggcgatgag	gcaggtcgcg	gcgagcgcg	gcgagcgcg	cgggacggcg	10380
gtccagccga	agagctgcgg	ttcgtagaag	aacgccagca	gcacgccgag	cagcatgaac	10440
gcgttcttga	accagtgatc	gatccgcgcg	atctggacgt	acggccaaat	ccgcgaaccg	10500
gagctagcgg	ctgacgacat	cggttagctc	ctcgagcgcg	cggtatcgca	accatccgcg	10560
gcccggctgtg	acgcccgtgg	ggccggcgat	gagggcgcg	gagagtcctg	cgcccatcgc	10620
tccttctccg	tccacgtccg	gtcgatcccc	gacatacagg	acctcccag	gtccgagtga	10680
ccacatctca	cacgccacat	ggaagcccg	cgggtgcggc	ttcagggcgt	tcacgttgct	10740
ggcggtcgtg	cagagcgcca	gcgaaaaatg	ctgcgacgtg	ccgagcgcg	cgagcttgct	10800
ctcgggcgcg	tagtccgaca	gcacgcccag	cctgagcccc	gccgcgcgaa	gcgagtcgag	10860
cgcgccgagc	aaacccgggc	gccggcacca	ccgcagatac	ttgagcgggc	ggcgaccat	10920
ccattcgttc	accgtcgccg	ccacgacctc	gttggtccagc	ccgagcgcc	gcgcggtg	10980
cgcgatctgg	cgcgcccgca	gcggctcgtc	ggccgcgcct	aggcgccgga	gatcctcatg	11040
cgccccggcg	tactcccgc	cgatgcgcgc	cgtgtccact	ccgcgcgcgc	accggaacgt	11100
cagcagcgcg	gtggcgcgca	gctccgctgc	catggcgggc	cgcaacggac	cctgccggta	11160
cagcgtgccc	tccacgtcga	acagcacggc	cttgaccccg	cgcaacggct	tcactctgcg	11220
aggcccagca	tccgcagaat	cgccgacacc	tttgacgcgt	cgcgaggaa	cacgtcacga	11280
accacacagg	cagccagccc	cgccaacagc	accgcgcgtg	ccgcgcccgc	tgccgtccgc	11340
cagctcccgc	gccaggcgcg	cgtgtcgatc	gcccgcggcg	cgtagcagag	cagaatcggg	11400
acgagcgcca	ggtgataccg	gggatgtccg	aagacgatcg	cgtgaatgcc	gctcatgaac	11460
acgatcaggc	tcaggagcag	cagggtgcgcg	cgccagttga	ccggccgcgc	gaacgagatg	11520
ccgatgacgc	cgagcaccat	gaccgcgcgc	tacgatccca	tgatggacac	cgccccggcg	11580
aagaacagcc	accgcggcg	acggtagaac	ccatttcgga	cgccggcgat	gaagtgcgcg	11640
tccagccccc	agaagtctgc	gaacttcagc	accgaacggc	gaagcgtcgt	tcccggatgg	11700
gccttcatgt	agtccagcgc	ctggcgctgc	gcccacttct	ctttcgtgcc	ctccgtccag	11760
cggtgcccgt	cgggcgcgcg	aggcgccatc	gtgcgcgacc	attccttgct	gcccgtgagc	11820
gagatcgctg	cccacatgcg	atcctcgggc	gtgtacgcgt	agttgcccac	catcaggttc	11880
agcccgcgga	gcgtgtccac	gaccgtgaac	gagcgttgca	ggagcgtgtt	gcgcacgctc	11940
cacgggcccga	cgatcgcggc	gtatccggcg	aagagaagca	cgccacgcgc	cagacgcgcg	12000
cgaagcggga	tgcccagacc	gacgagcgcc	gcccgcattga	ggatcgcgac	gtacggccac	12060
atgatcgagc	gaacgagcgc	ggacgcgcgc	agcgcgacgc	ccgtcgcgag	cgcgatccac	12120
gccgagtgcc	gccctgatgg	gcgatcgatg	agcgtcagac	atcccaagag	caacaaaaga	12180
aggagtgtga	tgaagagcgt	ctcgacagc	accagcacgc	cggagaacag	cagcgaggga	12240
tagaacgcga	acgcccccgc	ggcaatgagg	ccggcgcggtg	cgctgaacaa	cctgcgccc	12300
atgagatata	cgagccagag	accgagcagg	ctcaacggaa	cctgcgccc	ccggaccgcc	12360
gtcagcgatt	ccgagcccgt	gaccgcccag	acgcccgcga	cgaaacgggg	aaagagcggc	12420
gcccggatgg	acgtcggttc	ccccggcccc	cacgcgaacc	cgttcccggc	gacgacgttt	12480
ctggcaagcg	tcgcgtagtg	ctgctcgctg	cgaatctcga	gcgcgacgtc	cttgaacccc	12540
caaatcaggc	cgagccggag	gaccagggcc	agcgtgagga	tgacgagaat	tgcccggcg	12600
tcccaatggg	tgcgctcggc	ctcgtcgcgc	ccggggcgga	cgattggcca	ggggggcgac	12660
gatgggagtg	ggggagcctg	catcattcgc	taaaacaagc	agttacggtc	aaacctgtgc	12720
caccggggaa	tcagcaaaga	ctctgccaac	gggtccgcgc	gcggcgacgc	ccggcgcgcg	12780
ttttccggcc	ggaagtccgc	ccccaaaggc	accggcgctg	gcacagccca	gtagaaatct	12840
cattctggga	tgacttttgt	ccatgatcca	tgaccatttc	gaggggcctt	cgtgacgcc	12900
tggggatggc	ataagcttgg	cacctgcgat	ggcgacgtgt	cgatgatcag	cgggcggtgag	12960
gagctccaga	aggcctatcg	ggacgacggg	gtcgcccgtg	aatacgtggc	gcggcggttc	13020
cagtcaccgc	tgggggcgct	cctccattcg	cgtcagatcg	gggtcgtgcg	agagctcgtc	13080
cgcgcgagg	gcatccggcg	cgcgccggag	atcgaccccg	gtccagcgcg	gctgaccgtg	13140

gacatcgcgcg egctcctcga ccgcggtgacg ctcgctcgacg cgagcgccca gatgcttcac
gaagcgcgcg agcggtcgcg cgaccggacg ctcgccgcg gggccaatct cgtgcaggcc
gacgcggttc gtctgcccgt cagcgggcgg ttcgatctcg tctacacctt caggctcggtg
cgccatttcg agcgcgccga tcgggtcgga ctctaccgccc agattgcgac gattctcgga
ccccggcggt ggctcgtctt cgatgcggtg aatgagattg tgtctgcgcc gcttcgcgcc
aacgcgaagc ccggcgagta cgagcactac gatgctcgca ccgggtgcga tcgcccgttc
gacgagctgg ccgaggccgg gttcacgatg gtgtcgctca ccgggtgcga tcgcccgttc
cccacgctca tgaagtgcga gatgtatctc ggcggaggt caacgggtgct cgcgcgcgcg
gcatggagg tgatcgatcg gctcggcggg gaaccgctcg agtggatcgt cgtatgccgc
cgcggtgac ctactggacg ggtacctggg atccgatcaa agaagcgatt tcgaaggaaag
tgaacgccct ccgcaccggg tcgcccggcg acgcgacggt gtcgctggtt gtcggcgccg gcttgatcg
agcgacggcg tctcgtgccc aaagatcgcg tgctcggtt gtcgctggtt gtcggcgccg gcttgatcg
cccttcgggg cgcgcccgcc gtgctcgagc cgcgcgggcg cgtgacgcac gcttcggcg
gcggttctc ctggcatctc ctgagggcg tcggggcgcg gccggttctg ctaccgccc
tcaaggaaga tggcgctgcc gatctgccc gggcgatcgc gcacgtggcc gtcgaagtcg
gcccgcctcg cgacacgtgg atcgcgccg gtatgtctcc cgggtcggtt ctcgctggag agattcacgc
atccccgggt ggacctcgac tggatcggg ctgagatcga gccgcgcggc attccgctcg
tgctcttcgc cagcacgcgg tccgatccgg cgtccggaga ggtgcccgtg cgccgggtggg
tcgtcgagct ggcgcggtc cgcgcgctcg acgcccctcg tccgccaccg aacttcacg
gcgacgtcga cgctccagg cgcgcgctcg cctgttacg gcgggcccac gccaccgtcg
tctcgtacga ggacgcaccc gacatgcgcg gtcggaattt cgtgctcgag ggattggcca
tgtgtctcgc gggcgccacc ggcaaggcct acgcgatcgc ggagatcctc gatcgatcgg
tgggacggcc gtgtctctcg acgcgcgaga cgtcgctcgc cgcgagcctg gacacgctcc
gcgcggcgat catcgcccg cgcgatgcgc cgcgcgggt cgagaggag acgttcgacc
gcgcggcggt ggccggcggt gccgagcgcg cgtcgcggt cgagaggag gcccgcggtg
tcggccgctt ccgcgcgcgc tacgacgcgc tgtacgtgga ggtggcatcc gcccgcggtg
gcgagcacgc gaggcgcgcc tcttgacgaa gggactcgcc ggacggctgc tcgggtgggg
tcagaaggcg aacgatttcg ccgactacgt ccgcatcga tcgcagatga agcgcgacgc
gcggcggaat cgccgcaggc tccatcgggc gctcgcggg cgcggtgaacc ggtacctcat
cgtgcccgtg gccagcatcg gcgacgtcgc cctgacgtcg ggcgtcctcg gggccctcg
cgcgacgtac ccgcaggcga cgatcgactt cactctcag gatctgtcgc gctgtccga
ccggcaccac ccgcgatca acaccgtcta cgtgcccgg cgagccttc tcagaacgtc
gtacgactgg atcatcaacc tccagaacct cgtgcccgg cgagccttc ccaggatcat
cgggctctcg taccgcga ttctcgagca gctctcgagc cgagccttc tctactgcat
caccggccgt cactcgaga acggtcggga gtcacgcgc acgaacatcc cccgctaccc
ctgtgagatc gaggagcatt tctcacggc gctcgctccg gcaacatcga agttcgctt
ggtgaccgcg attcacgtgg acgcacagg cctcggggt aactcggtcg ggtgcccgc
cccggagaac cgccagggtg tggcgatctt cctcggtcc aactcggtcg ggtgcccgc
cgatgaggcg tttcgacgt actccatcga ctacgtgaag cgctcggtc gtcacttcga
agatcggttt gcgatcgcca tcatcgggca gtcccagggt cgcaaccgg cgaacgcgc
acagtaccgc gagatgctcg gccgccatcc ggagctgato gatctcggtg acaagacgtc
gctcgacgag ctcggtggcc tcatggactc gttttccctg ctcatcagtt ggcactcgag
ccccgtccat ctggcgatgg ccgcggagt gccggtcatc ggggtgtacg tcaacgacgg
gacgttcagg atgcacccga cgctcgccga cgagcgcttc gtggccatca acagcacgc
gccgtgcttc acctactcgt ggcgatggaa gttcttctgc tcgacgtgccc gcgaccggc
gaccagatcg cattactgtc acaacgagg gttcgtgttt agctcgatg caggtccgt
ggccgcaate gaccgcggcg tcgatcgaga cgaagcgcaa agcgtctctg ctcgctgatg
ttcgcgcgca gtttgagag ctctcatcg agcggtcga ggcggacac aagatcctcg
cggtggccga gtcggcgcg cccgcggact cgcagcattt cgcgcccag acgtcgatcc
tgtgcccga cggcggtcg gccgcggact ccgtggcgct caccaccgac tgagtgcctc
gcttcgagaa ggagcgccgc taccggtacg agaccgtgtc gcaactcgcc caacatcctc
tcacccggtg gccgaacgac catcggtatc acgacgtccg tgcgctcac cggacggcg
ggagcgccgg acgaggcgga aaagaaacgc atgaaacgaa tcgctccgctc agatcagacg
ggcggaaga tgaagagtct cgccggcgct tgccagggtc tgcctccgctc gatcgagaa
tcgaggatcc aggaagtaca cattaccgtc acgtcgatc ttcgaacgtc gaccaggagt
gcgtcagct gagctggcg tagacgtcga cgggtgagcg gccggcgagg gtcggcctgg
gctcgcggg gatgaggtcg ggcggcggt ttcggtagc cctcgcgctc ggtggagccg
cgaacaacag ttgcagtcca acttcacgca cgcccggtat ctcggtcggtt agaaagcgca
cggtcttccc gtcttcacg agcaccgaga gcgcgagccc gcggaagccg agaaagcgca
cgcgggccga ctgggcttcg agcagcaccg gcgctcctg cgctcgctcg gacatcagca
ggaaggcgtc ggggcctgg gagcgagag tggcgctcag caggcccgcg gacatcagca
cgcatcctc cctcggtcgg gtcgtcccg accagccag actcgaccgg cgtggtgag
gcctggtcgc ccggtattgc gtagcctgccc cttgtgtgag cgtggtgag cgtggtgag
ggacgacgaa cggtcggtc tcgagggcg cggccgtacg gaatcggtg gacggcacc
ccacgaactt ctcccgatcg acccggtcg cgatgtagcg cgcacacgtg acgatgtgct
gtccctgctg cgcccgctcg atcatcacgg cctcgggatc cgttctgatg tgaaagtga

13200 15/35
13260
13320
13380
13440
13500
13560
13620
13680
13740
13800
13860
13920
13980
14040
14100
14160
14220
14280
14340
14400
14460
14520
14580
14640
14700
14760
14820
14880
14940
15000
15060
15120
15180
15240
15300
15360
15420
15480
15540
15600
15660
15720
15780
15840
15900
15960
16020
16080
16140
16200
16260
16320
16380
16440
16500
16560
16620
16680
16740
16800
16860
16920
16980
17040
17100
17160
17220
17280

cgccacgcg	cgccgcgcg	gcctgcagcg	cggtctcat	gaatccgtac	acgacaggat	17340
tgtgcacgtg	gaccaggggc	cgcccacga	gacccaatcc	cggaagcatg	cgcgacacg	17400
cgccgagctg	ttgcaggccg	cgcccccga	gcgcgttcag	gtcatacatc	cgccgcccga	17460
cgccctccgc	gtcgagcgcg	gcgcccgcag	gaccctcgcc	tgtaatccat	gccgtgcacg	17520
ggacgcccgc	gcgtttggcg	gcggccgaca	accggatcgc	cagcaccgcc	gcgcccgcga	17580
ccagatgtga	gacgaggatc	tcgtgcaggc	ggggggactg	attcatggtg	gcgcgcgatg	17640
cgcccgggcc	gccagcctac	gttactcgcg	tcgcaggctcc	gcccgcgcgg	tagagcgagc	17700
gcagcaggtc	gcgaagtccc	gacgggagga	agcacacggc	gtaccgcgtg	agatcggtgc	17760
gccgcacgtg	caggccccgc	tgaccgaacc	gtttccgcgc	cgcgtagaac	tcgcccgcgt	17820
cgtactcgag	gtggccgagc	gcgagcagga	gacgatggcg	atgatcgctg	gcggcgcgat	17880
gcgcccactg	atcgctcccc	cgcccgagga	accgctccag	cgctcctgacc	atgccttcga	17940
gcacacgat	ggacgacttc	gacatgttgg	tgctgtgcag	cctgtaatgc	gtcaggggtt	18000
ccgggatgaa	accgaccggg	cccttcgccc	agaggcgag	ccagaggtcc	cagtccctga	18060
cgccgtggat	gccggctcgcg	aacatctcct	gccggagcac	ggcggctttg	accatgaccg	18120
acgaagcggg	cacggcggtt	gcgaggacga	ggtcggggag	acagcgcgcc	tcgagcggcg	18180
gaagatcgag	cgtagcgagg	acctcgccgg	cctcgtcgat	gacggcgcg	cccgtagaa	18240
tgaagctcca	gtcggcggtc	tgtaggaaga	gatcgacctg	ccggcgaaatc	ttggccggat	18300
cccacagagt	gtcggcggtc	agcagcgcca	cgaattcccc	tcgcccgcgc	gcgatggcgc	18360
gattgcgggc	gccggcaagg	ccccggttgg	cctgagagat	caccgtgatg	gcgcccgcga	18420
atctcgccag	gatgtctgcc	gtgccgtcgg	tcgatccgtc	gtcgacgacg	atcacttcga	18480
cgggctgata	ggtctgggag	agcgcggact	gcacggccct	ggcgatgtag	cgctcgccat	18540
tgaacaccgg	gacgatcacc	gacacgagga	cggggatatac	ggtctgggcg	gccggcgagg	18600
gctctcgagc	ggcttcagcg	gccatggccc	tctgcactgg	cgtagcgagc	gaacagcatg	18660
acgtttctcg	aaagccacgg	agacaacggc	gcgctcttca	tcgctccccg	cagcgtctgc	18720
gatgccgggc	cgaggtagct	gaaccccgct	gccgcgaaga	tcttctgcca	gtactcgagc	18780
ggctgttcgt	tgatgtggcc	cattccgccc	tggcccggct	gcgcccgcgt	gaagagcacc	18840
agcggcgccg	tgtagcgtg	cacggcgagc	agcttctcgc	caagcggcg	cggaagggtg	18900
tcggcgacat	cgaggctcat	ggagaggtcc	gttgctccg	cgcccggtga	caccgcggtg	18960
agatcgaatg	agttcagcgc	gacgcccagc	ttttccgcgg	cgtagccgtc	cggtactcgc	19020
gatcgctcgt	agccggccgc	cttcacgcct	cgtagcgagg	acgcgtgcac	gaagtgaccg	19080
gtgccgcagc	cgaagtgcgt	catcgatcgg	atgcggggtc	gcaacgccag	gaccccatca	19140
acagagcagg	gcgcccgcct	ggcggcgagc	tgcccgaaca	gatccatgac	ggccgcggtg	19200
tacgtccatc	gctcgaagcc	cagccgttcg	ccgatccggg	acatcttcag	tcggccctgg	19260
caacgccccg	accagtagag	ccgcttgagc	tgggtcacga	tcaggggttc	gtccccggcg	19320
cgacgcgaac	gcgcccgggc	gcagtcagcg	acgcccagaa	ctccacatcg	gccacgcccg	19380
cccacgatcg	atgttccctg	accgccagcg	tcgcaagctg	ccgcgacggc	acgaggaacg	19440
tctgggtcgg	gttgtagaac	acgtgcaccg	acaggtagta	gtggccgcga	acgagatgcg	19500
cgcgaaatg	gaggtcgagc	gagaactcgc	ccatgaacgc	ggacgcgtcc	ccgagatccc	19560
gcgggtggat	gttgccgtcg	tagacgacga	gctgatcggg	ggagcgttgc	accatgaagc	19620
cgaacgtgac	gtcgtccagc	ggctcgggtc	ttctcgcggt	gacctggagg	ttcagctgct	19680
tcgccggcgt	cacttcgctg	tgctgccccg	acgactcgtc	gcccaggaag	agcgagccgg	19740
acacgatctc	gacggggctc	ccgacactcg	gctgttcgga	tttctgggac	gcgcccagct	19800
acgcgtcgag	gacttcgatg	gccgtgccgg	cgcccttcac	cgagccggcc	agatacaccg	19860
catgctcgca	gagggctcgg	accgcctgca	ggttggtgca	gacgaggacg	atcgcgatgc	19920
cctggcgctt	gaactcctgc	atgcgatcga	agcacttctg	ctggaacgtc	atgtcgccga	19980
cgtcacaac	ctcgtcgatg	atcagcacgt	cgggccgcag	gtgcgcggcg	atcgagaagc	20040
cgagccgcgc	gttcgatccc	gacgagtagc	gcttgacggg	cgtgtccacg	aagtccggtga	20100
cgccggcga	ctcgacgatc	tcgtcgaacc	gctcggcgat	ctcgacccgt	ttcatcccca	20160
tgatcgacac	ctggaggaa	acgttctcgc	gcccgggtgag	gtcgggatgg	aaccccgcg	20220
cgacttcgat	gagcgcgcgc	gcgcccgcgc	ggacgatcga	gcggccgggtc	gtcggcttca	20280
ggatcttcgt	cagcaccttg	aggggtggtc	acttgcccgc	cccgtttgcg	ccgatgatgc	20340
cgagcgccct	gccgcgggtg	acttcgaacg	agacgtttct	gaccgcccag	aactcctggt	20400
cgtccagcgc	gtcggcgggg	cgcggttgaa	acagcccgcg	cacgagcgcc	ggcacgaggt	20460
cccgcaggct	gtcatgacgc	tcgcccgcgc	ggaacttctt	ggagacgttg	tcgaagacga	20520
cgggggccgt	cacgtcagac	gttctccgcg	aattcgaact	cgagcgtatg	gaagaacagc	20580
cagccgaaga	cgagcgtcac	gaccgagacg	atcgccgtga	cgcccagcat	cacggacggc	20640
ggctgggtta	ggagcagcac	ggcgcggaag	ccgtcagatga	tcggcgctcat	cggggttgagc	20700
tgagggccc	actggagctt	gcgcccacc	atttcgacgg	ggtagaccgt	cgacgacgcg	20760
aacatccaga	cggcgatgac	gacctcgaag	aggtatttca	cgtcgcggta	gaagaggttc	20820
gccatcgaca	ggatcagggc	gatggccgtg	gtgaatatca	cgagcacggc	gagcacggcc	20880
ggcagccaga	ggatgttcgg	tcggaccggc	acgtggtagt	agatcatgag	cgcgacgagc	20940
accacggagc	cgatcgcgaa	gtccacggcg	ctcacgatca	cgcgaggaaa	cggggaagatc	21000
tccttgggaa	agtacacctt	cgagacgaga	ttcgtgttcc	cggtcaggga	ggtcacggag	21060
aagcggagcg	ccgaggagaa	gaagtctccg	acccacaggc	cgcagaacgc	gaacaccgga	21120
tacgcgaccg	gcgtgtcgat	ggccgcgacg	cgcatgaaga	tcacggagaa	gacagccgtg	21180
ttgatgagcg	gcataaacac	ggcccacccg	aagcccatca	ccgactgctt	gtagcgagc	21240
agcaggtcgc	gcccgcgtcat	ctgcacagc	agctcgcgat	actcgaactg	ctcggcgagc	21300
atctcagggc	aatcgccag	gaaggcgcg	atcgtcagcg	gtgtcagtcg	cgccgctcgtg	21360
ccgggcccgg	catgagggca	ggtagcgctc	gagctgcccg	aagagcaact	gcacgaactc	21420

ggtcgcgcgc cgtcgcgcgc cgacgtcgcc gttttcgctc gacggcacgg ggctcgtcac
 gcgagcagag cggcggtccg atcggttcag ccgcacgcgc tcgtacacca tgtagatctt
 gctccagtag tgcgtggcga tcacgcgcgc gtggctctgg taccagtaca gcacgatttg
 ccgctcgatc cccttctgga tgacgtatcg gttgatctcg atgggctcgg tggggcccg
 cacttgacc agcgcgcggc cgggtgatac gggctcccag ccggcgcccg gcaggcagtt
 catcggcgag tggatcgtgt cgccttcgcg ctggctctgg tagtagccga tgtagagcga
 cagtagggc tgcgcgcggc ccaggtagat ccggttgatg tactcgtcca cgcgagcac
 cgccatgacg tccttggtga agggcgcgct gtgcgcgcgc gtcaccgct ccatctggaa
 cgggacgttc gcgagcggct ggcgaagcgg gacctgctcg gcgcgcacgc agcgcgcggc
 aaacgcgcgc gtgcgcgagga aacacacgct gagaatgatc agtcgtctgg tcatgaatct
 ggactcacac cgtggccggc ggggtcacgc cgggagcag aggcgtgaag cgcgcgagga
 ccagaccac gacgaacaac atgacgaacg cggcaagaaa gacgagccag cccgagaacg
 tgtggaagaa cccctgcgcc gcggcctcgc cgtagtagtg cgccgccacg cccgtgcccg
 ccacacgcgc gccgttcgcg ataatcgcg cgtggatcgt cgagagcgca atcgccaggc
 gcatgccgac gcgcggatcc gtgaagtagc cgtacacgat gcccagcgtg agcagcgaga
 tgagcgaccg gatcccgtg cagcctcgg ccacttcgag cgtcgtgtgc gcgagcacga
 tcacgttgcc ttcgcgcagc accgggatgc tgagcgcgga cagcacgccc tcgcccacgc
 gagacgccag gagctgcagc ggaatgcca tctgattgaa gatgatcgac gggatcgga
 tcatcagcac caggaacgcg agcggaaacg cgagcaccga gagcttgccg cagccgagca
 ggaacacgat cgcgcggta agcggagcta gtatcgacac gcgcgtcaga aagagctcgg
 cgccgagcaa ccccgcgacg agcagcgcg gcccgcgag ccccgaggca aagtaggccc
 cgtcgcgcgt gtccggcagc gccgagagac gcccgcggcg cttccaggca gttgatccag
 cgagcggtac gatcagaatc ccgtgggagt agttgtcgtc gttgatccag tccgtcgcca
 accgcacgat cactgccc taaaggagca ggaactcac gggcgcgata ccgagcggaa
 ccagtctggc ctgaagcaca ctcatagcga tcgggacgaa tatagcaaaag ggctctccc
 ggcagaactt gggcgccgaa acacgtaacc atcacgtca gaatctaccc aatggatgac
 gacgggattg gccggtgga ggggctcgt gaagatcatc ggcctcttca gatgcaggac
 atcaactcct ccttttgact gcgtagcagc gaagatcatc cgccggtcaa gtgaaatccc
 aaatgtcatc tgtcaggaag agtaatgtct ctggccacag atccggcgag tgcgcggccc
 ccgacagaat cgcgtaccac cgcgactggc atccggcgag gttttcgctc agtatcttgg
 gaaaactgca ttacaggaac gccaggtcgg cagcggggtc gcgaccgggc cgatcacgcg
 ggcattgagc gcgctttcgt ctgcggggg gcggaaccgt agcccggttc gggagacggc
 tggagcaggg cagggacagg catggatggg tccgtgagac agcccggttc gggagacggc
 cgggctcggc gggagctcgt ccacctggcc gtggaccgcg ctctcgtatc gcggcctgac
 gcgcggcgcc taagttttga cggcgtctcg ggcgtcgtatc cgtgagtggt ggtcggcatc
 aacgcgctcg cgcacgcct gcagcgtcat ggcgtcgtatc ccgtgctcaa ggcggcgggc
 tccctcgacc ggtccatcga gctcatcgtc gggatcctcg aagggaccgc tgcagtagat
 gcctacgcgc cgtcgtatcc tgcgtatccg agggagcgc gtcgtatccg cgtcgtggg
 gccggggcgc ggatcgtgct cactcgcga cgcgtgagcg cccggggcgt tggcggaaga
 gtcgagctca tctgcctcga tgcggacgag cgcgtacgca tctacacctc gggatcgacc
 ccggacgcgc gcgtcgatct egaccacctc cgtccgttgg cgaatctcat cggctggcag
 gggcgcccca agggcgtcgc gatgccgcat cgccggcgcg cgcacgctcc agttcacttc
 gtcgagcgc cggcgtcgg cgccggcgcg cgcacgctcc agttcacttc gggcgagct
 gacgtctcgt tccaggaat ctttgcgacg ctgcgcgacc ggcgcgagct tgcgtcgtg
 tccgaggaca cgcgcggga tccgcgcgc ctgctgcggg tgctgcgcga tgcgtcgtc
 gcccggtgt atctcccgtt cgtcgcgttg cagcagcttg ccgtcaccgg caccgacgat
 ccggatctgc cggcgttcgc tgaggtcatc acggccggcg agcagctccg ctgcagccc
 gcgtcgtcc agtgggtcgc gggcatccg cagtgcacgc cccacaacca ctacggccc
 gccgaagcgc atgtcgtcac tgcgcacag ctgaactagc ccccgcgga atggccggcc
 ctgcgcgcga ttggcgtgcc gctgccgggc gtcgagatc tgatcgtcga cagcatcgc
 cagcccgctc cgcacggcaa cgtgggcgag atcctcattg gggcggtg tctcgcaccc
 ggctacttga atcagccgca gcgcacggcc ggcgtgtg ggcacgatgg cgcgacgatg
 gatgctcgcg tgtatcgac cggcgatctc attcgaggct tccgtgtgga gccgggcgaa
 ttctcgggc ggacggacga ccaggtaga gttcgggaga ccgcgctcgt cgcgcacgac
 atcgagaccg tccctcttca gcatcccggt tatgtgattc cgtcagccga tcaggcgggc
 gacgcgtcgg gcagcaagag gctggtcgcg acaggctccc ggtggcaga ccgtgtggga
 accgagcgc agggcttcgc ggatccggcg cttgcccgga cgggctggcg cagatgagac
 cgcgcggcg cggactcgac ggcggagatg cgggagtggt gagatcggt gggcacggg
 acggggatgc cctacaccgc atgcccacg cgtgtctat agcgcgacgg ggcgcagcc
 atccggtcgc cgcgccactg gggcccacgc gacggcctgc cgttccggcg gccgcgttcg
 ttccgtctcg gggcccacgc cgttccggcg cgtggactat ctgacgcgcg tccctcgcgg
 gaccacgtcc gggcccacgc cgttccggcg cgtggactat ctgacgcgcg tccctcgcgg
 gaggcgcggc actttaccgg cgttccggcg cgtggactat ctgacgcgcg tccctcgcgg
 gtcgcgcagc acttgccgtg cgtggactat ctgacgcgcg cgaacctgcg cctgctcga
 accgtgcgtc cggggcgatt catcttcgtc ggcgacgtgc atcttcgat tccgatgtt
 atgtttcacg cgtcggtcga gctcacgaaa gcgcggcg gacgcgagc tcgtgatcga
 cgcgagcgc tgcagcgtca gatggcgctc atctcgggag cccaggtcca gatccggcag
 ttcgacgcgc tcatggaccg catccccggc aggtacgacg tgcgtgctcac ggtggcgga

21540
 21600
 21660
 21720
 21780
 21840
 21900
 21960
 22020
 22080
 22140
 22200
 22260
 22320
 22380
 22440
 22500
 22560
 22620
 22680
 22740
 22800
 22860
 22920
 22980
 23040
 23100
 23160
 23220
 23280
 23340
 23400
 23460
 23520
 23580
 23640
 23700
 23760
 23820
 23880
 23940
 24000
 24060
 24120
 24180
 24240
 24300
 24360
 24420
 24480
 24540
 24600
 24660
 24720
 24780
 24840
 24900
 24960
 25020
 25080
 25140
 25200
 25260
 25320
 25380
 25440
 25500
 25560

cggccgggccc ggcgcgcgtcc gtccgctcgcg ctccgactggc gtaccggagcc ggtcgcgcgag
cgctcagctcg ggcaactgct cagcgccgat gcgcgggacg cgatcgagat cagggaacgtg
accaacgcgc ggctgctca gccggccgcg tgcttcgacg tcgtcaagcg tgccgactgg
accggcacgg ccggcgagac gcttgagagg gccgcgcgag acacgcgcga cgcgatcgac
ccggaatcgt ggccgctct cgcgaacgcc gccggctacg acgcccagtt cacctggtcg
gccgagatcg ggcgcacggt cttcgatgcg cgggttcgagc gccggggcac cgcgtcgctc
gccgctccc tgactcgtca cgcggttccc gccgcggagc gccggggccat tcagcagtac
gcgacgaatc cgtccgcga tgcggtggcg cggcgccctca cgcgcgagct tcggcgattc
ctgcaggagc gcggtccgga tcacctggtg ccgtcggcct tcgtcgtact cgtgcgatg
ccgctcacgc caagcgggaa gatcgatcgg cgcgccttc cggcgcccga gagccgcgt
cccgatctgg acgtggcggt tgccaagccc gccacggagc tcgaacgcaa gatcgcgag
gtctggcaga cgacgctcca gatctcgtcg gtcgggattgc acgacaactt cttcgatctg
ggcgggcact cgtcctgct cgcgcaggcg ttcgagcgta ttcggccgct cgtgcccggc
aagcagtggt cgatgatcga gatgttccag taccgcagcg tctattcgct cgcggggtt
ctcagtgagc ggccggagcg cggcgccggg cagcttgccg gcgcgcagga tcgcggacgc
cgccagcgcg agatgctgac gcgctcaggg cctgcccgtga aggggtccgg
taatgcagaa cggcatcgcc ctcgctcgga tggccggccg ctttccgggc gcgcggtcga
tcgaggagta ctggcgcaac ctcgctcgcc gcacggagtc gatcacggtc ttcaccaacg
aggagctggc gagggccggg atcagccaga gcgatctgca gaatcccgc tacgtcccg
cgcgcgggct cgtcccggat ccggacctct tcgacgccc gttcttcggc atcagccagc
gcgaagccga gctgatggat ccgcagcacc ggtcttctc cgaggcggtg tggctcgcg
tcgaacacgc ggcatcgcg ccgcggctcg tgcgtcgaa tctccacagc cagcgggat
tgagcacggg gatgacgaac agcacgtacc tgcgcggcgc tgctcgga cgagaacgac
tgctccagcc gtacaagctg cacttgccgc gaccgagcat gaacgtgcag accgcgtgct
gccgcgtctc cgtcgcggg gtgcaggcg gtacgtcgct gctcacgtgg cagtgtgatg
cgacgtcgct cggcgcggtg tccgtgctcg tcccgaaca ggaaggctac gtctacgtcg
cggcgctcgc cggcgcggtg gacgggcact gcccgcggt tctcggaag gctcgaggac
aggcgggcat cgggtcgcca ggcctcgtcg gatcggcgct gaacaacgac ggcctcggcga
cggctctcag caacggcgct ggcctcgtcg gatcggcgct ggtcgtggcc atggcgaga
acgggcagac gatctacgcg cgcagccccc agcgtggagc gccagcgga ggcgcacggc
aggtcagttt cgtggcgccg cagacgatcg actacgtcga ggcgcacggc ctttcgcgc
gcgcgcgga cgtggcgccg gtggctgccc tcacgaaggc tttcgggcac ctcgactcgg
cgtcggaga tccgatcgag ctccggtcgg tcaagagcaa ccatcaccag cagattccgg
cggagaacgg gttctgcgga atcaaaacgg cgtcgcgct cgccaacagt ccgttctc
cggccggcgt ctagctcgag cccaacccga agtggtattt ggcacgcgcg gccgtgagct
cgacgtgca gctgcgacca tggccgcgc tcgtgctcga ggaagcgccg cccgcgcg
tgaacgcgc cgttcgggat cggcgggacc gccagcgacc gtcctcga ggcgcacccg
agcgaccggg cgcgggctgc aacctcgcga gtcctcga ggttcgcat cgacgagcgc
cactcgacgc ggcgcacacg ctgcgcgcg gccagcgga ggttcgcat cgacgagcgc
tcgcgcagct cagccgcacc gatcgatcg cgtggttcg cggcgcgga gccaacgtg
tcgctgcac cagccgcacc gatcgatcg cgtggttcg cttcatgttc tccggcgcg
tcgccaccgg cgacgcgaa ggcgggag tgcgtccggt cgcgcaggtc ttcgcgaag
gcacgcagta cgcgaacatg ggcgggag cctccgcga cgtccgcgt gcgtgtttc
aaaccgagcg ctgcctcacg cggcccgcg cctccgcga agctccgtc gttggtttg
cggatcatgc ggcagtcgag cggcccgcg cccgtcagtg gatggcggtg ggtgtgcagc
ccgcgctctt tgcgtgcag tacgcgtcg agcatggcg agtacacggc cgcgtgcctt
ccgacgggct catcggtcac ggcgtcgtg gtctgagagg ccggtgttc gagacgctt
tgtcgtcga gaacgcgctc gtcgccgtt cagaggcaga ggtgcggccc gcgtcgcg
cggccgggga gatgctcggc gtcgccgtt caagctcgt cgtcgtttcc ggcgcgcgcg
atggactcgc gatcgcgcg gtcgtgatg cggagggcg tgcgaatgc cgcgcctcc
acgcgctcgc ggcctttgc cgcgtgatg tcgatccgat cctggcgag ttcgagcgtt
acatcgcggt cgcggcgac catcgcgct gacgcgcaa cgtgtcgaac ctcagcggt
tctccgcgg cgcggacgaa gcgaggtcgc cgcagtcga gacgcggcat ctgagacaa
cgtggctgac cgcggacgaa tgcgcgca tgcgcgca gccaatcgg gtgctcctc
ccgtcaggtt cttcgacggg ctccgacgc tcgcacggc ccatccacag aaaacggcg
aagtggggc cggacagag ctgtccggat ccaaggaacc ggcgcggag gctgtgttc
cgcacgtcgt tcaagcgacg ctccggcag caggggtgc tccggtcga tccggtcga
tgacagagg ggtgggcagg ctctggacg cgtgtgcgc tccggtcga aagaagcgt
acagaggcga gggcaggcgt cgcgcggtc acgacaaact tccggtcga tccggtcga
acctggtcga atcgccccg acgacaact tgcacaaact gagcgccct cccgagggc
ccgatgctac tcagcagatg accgaaatc tgggattcga ctcggtgttc ctgacgagg
atcgtatcc cgcgctgtc tttctggaga gagttcaagg gttccgcaag ctttcgaag
agatcgatcc gttcaaggcg gacgtcagc acatcgagc caagctgcct gcggacggt
cgagctgcg gtcacgagc cttcgtggc cgcaagcgt cgttcgcgg
aggcaccgac ggcggaagc ggcgtgcgg cggcgcgcc cgcacgctg cagcacgtg
ttcctgcgc ggcggaagc ggcgtgcgg aactccgcat gctggggga aatccgcgc
gccctcatgc ggcctgggt atggccgagc

25620 18/35

25680
25740
25800
25860
25920
25980
26040
26100
26160
26220
26280
26340
26400
26460
26520
26580
26640
26700
26760
26820
26880
26940
27000
27060
27120
27180
27240
27300
27360
27420
27480
27540
27600
27660
27720
27780
27840
27900
27960
28020
28080
28140
28200
28260
28320
28380
28440
28500
28560
28620
28680
28740
28800
28860
28920
28980
29040
29100
29160
29220
29280
29340
29400
29460
29520
29580
29640
29700

tgctgcccgc	gatgcccgcg	ccagcgtcgt	ccgtcgccgc	ggccgacgcg	ggactgaagt
cccgcgctac	gacggcgccc	gtctcgaagc	cgcctgcggc	cttcaagccg	atcgacggca
aagccagtg	ggagctgtcg	cccagcaga	tgccgcgct	cgccgcctgg	atcgatcgat
acacacggcg	cacggccgga	tccaagcgac	gcctggccga	gtaccggccg	gtgcttgccg
atccacgctc	ggcgccgga	ttcaaccggc	tctggaagga	gatgatctac	ccgatcatca
ccgaccgggc	ggaaggatcg	aaggtctggg	acgtcgacgg	caacgagtac	gtggatctcg
tggggcggtt	cggctcgctt	ctgttcggac	accgttcgcc	gatcgtggtc	gacggcgtgg
aggaacaact	ccaccggggg	ttcgagctcg	gaccgttgcc	gcccgtcgcc	ggcgaggtcg
cggcgctcgt	caagggaattc	acgggggatgg	agcgcgttgg	attctccaac	accggctccg
aagccgtgct	ggcgccacg	cgttttgccc	gaacggtcac	cggcgcgac	aagatcgccg
tcttcgaagg	gtcgtatcac	ggtctccttg	acgaagtgtc	cagccggccg	ctcgtggtga
acggcggaatt	gcggtcggcc	cccgcggcgc	ccggcatcgc	cggcagcgcg	gtgagcgaag
tcacctcgtc	cgagtacgcg	aacccgaagt	cgctcgaggt	catccgcgcg	cacgagtcgg
agatcgccgc	gatcctcgtc	gagcccggtc	agagccggcg	gctcgacctg	cagccgaagg
agttcctgca	cgaggttcgc	cgcgtggcgg	acgagatcgg	cgccgcgctg	atcttcgacg
aagtcatcac	cggcttcgcg	gtgcatcccg	gcgccgccc	ggcgcacttc	ggcgctcccg
cggatcttgc	cgcctacggc	aaggtcgtcg	gcgccgcat	gcccacggc	atcgtcgccg
gcaatgcgaa	gttcatggac	gcgttcgacg	gcgccactg	ggagtacggc	gacgggtcgt
tcccggaggt	gggggtcacg	tactttggcg	gcacgttcat	gcggcatccg	ctcgcgctcg
ccgctgcgaa	ggccgtgctg	acgtacatga	agtcgcaggg	tcccggcctg	cagacgcggc
tggcagatcg	gatcgagcgg	ctcgcgaacg	acgcgcgcgc	gatcgtggcc	aggcacggcg
cgccgatatca	gattacgcag	ttcagctcga	tgtagacct	gaacttcccg	cacgaccaga
agacggcggt	gctgctgttc	ttcctgatgc	gcgagcgcg	gattcatatc	tgggagggcc
gtcccttctt	cttcacgacg	gctcacaccg	aggccgacta	cgccgccatc	ctccgcgcgc
tcgacgagtc	gctggccgag	atgcaggcgg	ctggcttctt	cggcgcgccc	gcccactccg
cattgcaggt	gaccggggacg	gcgaccagg	acgtcctgcc	gttcaccgaa	ggacagcagg
agatctggct	cggcgcccg	atgggcgagg	cggcatcacg	cgcctacaac	gaagtctcgc
cgtcgacct	gcgcgggcct	ctgaatcgcg	cggcgatgca	gcgcgcgctc	gacgaggtcg
tcgcgcgtca	cgaatcgctg	cgcgatgacg	tggccagcga	accgctcggc	ctgctcacg
ttccgggcac	gtcgtggtcg	gttgaatggc	aggacgtgtc	gggattggcg	gaggacgctc
agcacgaggg	cgtgcagaag	gtgctcgagg	gcggcgacgg	cgtggcggtt	gacgtggaa
gcggggcggt	gctgcgcgtg	accaccatga	cgtcaccccc	cgagcatcac	gtcttcgtca
tggcgggccca	ccatctcgtg	tgtgacggct	ggtcgttcgg	cgtcatcctg	cgcgatctgg
cgacgttcta	ttcggcccg	gttcgaggca	cgcgtgcgaa	cctcgaggcg	ccgatgcagg
tgagccgctt	cgcgcgtgag	gatcgcgagg	ccaagcagag	tctgaagcg	gcccagaccg
aggcgttctg	gatccggcaa	ttcgacaccg	tcccagagcc	gctcgagtta	cccgcgcatc
ggccgcggcc	gccagtgaag	tgcgtaccag	gcgcgcgtgt	gtccgtgccc	tttgacgcgg
cgtcgcgtcg	cggcgccaa	aagctcgccg	ccgaacaccg	cacgaccctg	tttacgacgc
tgctcggcgt	cttcgagacg	ctcgtctacc	gactgaccgg	ccagcaggat	ttcgtcgtcg
gcatcccggg	cgcgggtgcag	cctctgctcg	gcgaggatct	ggtcgcgcac	tgctcaact
tcgtgcccgt	tcgcgcgcgc	gtgtcaccca	ccgcgacgtt	ctcgagtagc	ctgaacacgc
tgaggacgca	gtcactcgac	ctcaatgagc	accagaactt	cacgtacggc	agtttgctcg
ccaagctcgc	gttgccgaag	caccggggcc	gcgatccgct	cgtctcgacg	agcttcacgc
tcgagcccg	catcgtcgat	cccgggttcg	acggctctac	ggcgcggttcg	ctgacgggtc
cgcgcgtgac	atcgaagcgc	gacctgcacg	tgaacgtcat	ggagatggac	ggcggtctgc
tggctgaggg	cgactacagc	accgatctgt	tcgacgaacc	gacgggtgcg	cgggtggtcg
atagctatcg	gatcctgctc	gaggccgtcg	tgtcgtcccc	cggcggtcc	ctcctgagcg
tgccgggtgat	ctcggagccg	gaccggagtc	agcttgtcac	cggctggaac	gacacggccg
ccgactaccc	gcgcgatcgc	gtcatccatc	aactcgttga	agagcgcgcc	gtgcagacc
cggagggact	cgcgctcgtc	tgtggtgccc	agcggctgac	ctggaccgag	ctcaaccgac
gcgcaaaccg	gttggcgcat	gcgctcgtga	agaaggcggt	cgcggcgggg	agcagggtgg
cgtcgtgcct	ggaacgatcg	gcccgggttca	tcgtctcggg	gcttgccgtg	cacaaggccg
gcgcgcgata	tgtaccactc	gaccccgctc	cgcccgccga	tcgcaagtc	ttcgtcgtcg
aggattccgg	tgccgtgctg	gtcgtgacgg	attcgcgggtc	cgaaggccga	tggtgacgc
ccaggacgcc	cgtgctccac	ctcgacgcgg	acagtcgcgcg	aatcgcgcaa	gagtcgcacg
atgatctgaa	ggtcgcctcg	tcggccgaag	acgcgcgcta	cgtgatctac	acgtccggat
cgacggggaca	accgaagggc	gtcgtcatcg	ggcacccgca	gctcgtgaat	tacacgtggg
cgggtgatcga	acggatcggg	ttaccgcgg	gctcgagtta	tgcgctcgtg	tcgaacgtcg
cggccgatag	cgggtgac	gggtctcctg	cgtcgcgcat	gggctgggtg	ctgcacgtga
tcacctcaga	ggctcgac	gacggggcgg	cgtcgggttc	gtatttgacc	gttcacgaca
tcgacttctc	gaagatcacg	ccgtcgtcgc	tcgcgtcgtc	gctcggcgat	caccgctcgg
cctcgtcgtc	gcccgcgcgc	tgtctcatgt	tgggcggcga	gcccgtcacgt	ccggcggtggg
tggaagagct	gcagcgtctc	gcaccaggct	gcccggatcat	gaaccactac	ggcccgcagc
agaccacggg	cggcgtgttg	acgtactgga	ccgggtgaccc	cagggatctg	ttcgtcgaga
gcgtgcgggc	cggcaggccg	ctcgggaacg	tccgcgccta	cgtgctcgat	ccggccggag
cgcggcgccc	gattggcgctc	gtcggcgagc	tctgtatcgg	cggcgcgctg	gtggcacgcg
ggtatctcaa	tcgtcccgcg	ctcaccgcgg	agcgtctcgt	ccccgatccg	ttcgccaccg
agccggggagc	gcgcgtgtac	cggaccggcg	accgcggccg	gttcggggcc	gacgggaaca
tcgagttcct	cgggtcgcgcg	gactatcagg	taaaagtacg	cggattccgc	gtcgagctcg

29820
29880
29940
30000
30060
30120
30180
30240
30300
30360
30420
30480
30540
30600
30660
30720
30780
30840
30900
30960
31020
31080
31140
31200
31260
31320
31380
31440
31500
31560
31620
31680
31740
31800
31860
31920
31980
32040
32100
32160
32220
32280
32340
32400
32460
32520
32580
32640
32700
32760
32820
32880
32940
33000
33060
33120
33180
33240
33300
33360
33420
33480
33540
33600
33660
33720
33780
33840

gcgaaatcga	agcggcgctt	cgcgctcatg	agggcgctga	gcaggccgta	gtcgtgctgc	33900
gaaaagatca	gcccggcgac	gagcgccctga	tcggctatgt	cacgaccggc	ggcggcgcc	33960
cgatctcgat	ggcggagctt	cgacgctcgc	tcaagcagag	gttgccggac	tacatggtgc	34020
cggcggtgat	cgctcgcgctc	gacaagctgc	cggtcacgtc	gattggcaag	atcgatcgac	34080
gcgccctgcc	ggcgccgccc	gagcgcgcca	gtttcgagtc	ggagttcgtc	acgcccgcga	34140
ccgagacgga	gcgcgggctt	gcggagatct	ggagcgcggt	gctcggtctc	gagaagattg	34200
gcgcgcacga	caacttcttc	gacctgggcg	ggcactcgct	gctcgcgggc	cgcacccgcca	34260
tgaaggtccg	cgacacgttt	cgctgggacg	tgctgctcat	cacgttcttc	gaaatgcccc	34320
ccatcgccat	gatggcgcca	gcaatcgacg	ccggtaccgc	ggccctgccc	tcccggacca	34380
tcacgcccag	ggcccgcgctc	agggcccgcga	gggatgagct	ggtccgcccc	tgaaccccgg	34440
ccacccgcag	actgagctcc	ccgacctgcg	gcgcgacggg	atcgccgccc	aagcgggcga	34500
cgtgtttctg	gcgcggcgct	cgtagcccca	ggggcgcttg	tggttctctg	cgtagctcga	34560
gcctgaaagc	tcgcgctaca	acattccgat	cgcgctcaga	ctgaagggcc	gcctcgaccc	34620
cgcgcgctc	gaacgcgcca	tcaacaccat	cgtccagcgt	cacgagacgc	tgccgacgac	34680
gttctcgctg	cagggcgatg	agctcaagca	ggtcatctct	ccgacgctgc	gtattccgat	34740
ctccctcgctg	gacatgcgctc	tgctcagtc	tgagaaacgc	gaggccggat	ggcgcccgct	34800
tgcggccgac	gacggcgcg	gcgtcttcga	tctggctgag	ggcccgctgc	tcaaggtcgt	34860
actcgctcac	ctcgagccct	ccgatttcgt	octcctgac	accgtgcac	acatcgtctt	34920
cgcgggttg	tcggcgggga	tcttcacccg	cgagctggcc	gaggcgtagc	cggtgagct	34980
cgacggggcg	gacccgaagc	tgccggagct	gccgatccag	tacgcggact	ttgcccgctt	35040
ccagaaggaa	cagaccgaag	ggcccggcct	gccggtgctg	ttgaagcact	ggacgcagga	35100
gctggcgggg	gcgtcgatga	cgctcgacct	gccagcggtg	cgctccgctc	cgccctcgca	35160
gacctcccg	ggcgccctcg	cctcgcttcg	gctcgacgga	gcgatctggc	cgggcgctcg	35220
cgcattcagc	cgggcgcgaga	acgcgacggt	gttcatgacg	cttctctccg	ccttcggcgc	35280
cgtgctgcat	cgctggctcc	gacaggaaga	catcctcatc	ggcagcccg	cgccggcccg	35340
cgatcgctcc	gagcttgagg	gcctgatcgg	gttcttcatc	aacacgttcg	tcctgcggct	35400
caacgcgctg	ggagatccca	cgtttcgcga	gttcgctcgg	cgggcgcgcc	gcgcgtgccc	35460
tgccgctcac	gcgcacaggg	acctgcccgt	cgatcggtc	gtcgagacgc	tcaacccgga	35520
gcggaccg	gatcggcac	cgatctatca	agtcatgttc	gcgcaggatc	cgccggcgct	35580
gaagatgacc	ttcgccggga	tcgagctcga	gcggcttgct	gtccacaacc	gcagcaccac	35640
gtgcgacctc	acgctcgagg	ttgccgaaga	cgccgacggc	gtcaccctgt	acgtcgagta	35700
cagcgtcgac	ctgttcgatg	cggcgacggt	ggcgctcgctg	ctgcagcaat	atcgccgtgt	35760
gcttcagacc	gcgatcgcg	acccggacca	gcggctctcc	gagctgcgtc	tgcttgccga	35820
cgacgagcgc	atgcagctga	tcgcactggg	gaccggtccg	gcggccgctc	acccgcgcga	35880
cgtcagctcg	gccacggtct	tcgaagagcg	ggtccgcgcg	actccggcgg	cagtggcccg	35940
gacgctggag	ggacagcacc	tcacgtacgc	ggatctcaac	cggcaggcga	accgcctggc	36000
ccggcggttg	aaggcggtcg	gtgtggggccc	gaacgtcgctg	gtcggcggtg	cgctcgaccg	36060
gtcgttcgat	ctgctgggtg	cgctcctcgc	cggtggcgaag	gcggcgcgcg	cgatctctcc	36120
gctcgatttc	gctcaccgcg	aggaacggct	tgcttcatg	ctcgccgaca	cgcgcgcgcc	36180
catcgctgctc	acgctgcggc	gatttgccgg	ggcgctcgag	tcctttcccc	tcttctgcct	36240
ggacgacgag	ctgcccggcca	ccgcccagga	gccggacgaa	gatctgccgc	cgcagtcgac	36300
cggagaccac	ctggcctatg	tgatgtacac	ctcggtctcg	actggtcagc	ccaaggcgct	36360
cgcgatcacg	catcgcgggc	tggttcgtct	cgtacgcggc	acagactacg	tccgctggaa	36420
cgacgtgcgg	gcgatcctcc	acatcgcgcc	gacgtcgctc	gacgcgttga	cgcttcgaggt	36480
gtggggcgcg	ttgctgaacg	gcgcgcggct	ggtgctgggt	cccagcgggc	tcgtcagcct	36540
ggagacgctc	gagcgcacgc	tcgggtcgga	gcaggtggac	tgctgctgct	tcacgaccgc	36600
gctcttcaac	gccgtcatcg	acgcgaagcc	cggcattctc	gtcggcggtga	agcaactcct	36660
gatcggcgcc	gaagcgctgt	cagtcgcgca	cgtgcggcgc	gccctggcgc	acctgccgga	36720
cacgcggctc	gtgaacgcgt	acgggtccgac	cgagtgcacg	accatcgcg	gcgcctacca	36780
gattccgcgc	acgctcgatc	cgcaggcgcg	gtccatttcg	atcgcccggc	cgattgccaa	36840
tacgcaggcg	ctcattctcg	atcgacacgg	cgacctcgctg	ccgttcggca	tcgcggcgca	36900
gctctgcctc	ggcggagacg	ggctggcgcg	ggaataacctg	aatcagccgg	cgctcactgc	36960
ggaacggttc	gtgcccacg	cgttcagccc	cggagccggc	gcccggctgt	accgcacggg	37020
agacctcgcg	cggatgcgcc	gcgacggcaa	catcgagttc	cttggccgga	tggtatcgcca	37080
gatcaactg	cgcggcttcc	ggatcgagcc	gggcgagatc	gagcatgccc	ttcgtgcccc	37140
gggagacggt	ctggacgcgg	tggtggagat	ccgggtcgat	gcccacatgg	cccatctcgt	37200
gcctacgctc	gtgcgcgccc	atgggagtca	actcaccgga	accgacctgc	gcgaacgtct	37260
gaaggcgcg	ctgcccggat	actgtgtgcc	cgcgaagtac	gtgttctctg	ccgaggtccc	37320
gcgaacggcg	gccggcaagg	tcgatcgctc	ggcgctgcct	gcacggcgga	tcgaggcgcc	37380
cgcgcccag	ccgcgcgatc	cgctgctcct	gaacgaggctc	gaagcggagc	tcgcgcgaat	37440
ctggagcgaa	gtgcttcggc	tgccctccgt	gctcgcgacc	gacaattttt	tcgatcgtgg	37500

21 /35
Figure 5b

CDS	complement(3..914) /blast_score=2e-66 /blastp_match="AE004644.PA2177 Pseudomonas aeruginosa" /gene="regulator hybrid" /note="probable similarity to prokaryote sensory transduction proteins" /product="sensor/response regulator hybrid"
CDS	924..2168 /note="none" /gene="ligase" /blastp_match="AP003013.MLR8297 Mesorhizobium loti" /blast_score=e-152 /product="2-amino-3-ketobutyrate CoA ligase"
CDS	2207..3190 /blast_score=e-151 /blastp_match="AE008872.TDH Salmonella typhimurium" /note="none" /gene="dehydrogenase" /product="threonine 3-dehydrogenase" /pfam_match="PF00107; adh_zinc; 1"
CDS	3373..4455 /note="putative" /gene="methyltransferase" /blastp_match="AE001866.DR0026 Deinococcus radiodurans" /blast_score=1e-08
CDS	4546..4959 /blast_score=2e-11 /blastp_match="AP002997.MLL1617 Mesorhizobium loti" /gene="unknown" /note="pfam00263, GSPII_III, Bacterial type II and III secretion system protein, Expect = 7.8"
CDS	5176..6192 /blast_score=5e-98 /blastp_match="AF064070.PE20 Burkholderia pseudomallei" /gene="glucose epimerase" /note="putative " /product="UDP-glucose 4-epimerase" /pfam_match="PF01370; Epimerase; 1"
CDS	6331..14043 /note="substrat AT, malonyl ; zinc depend dehydrogenase ; zinc dependent adenosine deaminase putative" /gene="PKS I" /blastp_match="AF285636.WCBR 2547 Burkholderia mallei" /blast_score=0.0 /pfam_match="ketoacyl-synt" /pfam_match="ketoacyl-synt_C" /pfam_match="Acyl_transf" /pfam_match="SAM binding" /pfam_match="adh_zinc" /pfam_match="pp-binding"
CDS	14275..15408 /blast_score=e-104 /blastp_match="AF285636.WCBT Burkholderia mallei" /gene="acyl-CoA transferase WcbT"

CDS
 /note="putative"
 /pfam_match="PF00155; aminotran_1_2; 1"
 15436..16245
 /blast_score=5e-11
 /blastp_match="TTDEFFMT.FMT T.thermophilus"
 /gene="formyltransferase"
 /note="evidence experimental"
 /product="methionyl-tRNA formyltransferase"
 /pfam_match="PF00551; formyl_transf; 1"
 /pfam_match="PF02911; formyl_transf_C; 1"
 CDS
 16287..17384
 /note="putative"
 /gene="glycotransferase"
 /blastp_match="AF285636.WCBD Burkholderia mallei"
 /blast_score=e-99
 CDS
 17427..18158
 /blast_score=2e-82
 /blastp_match="AF285636.WZM Burkholderia mallei"
 /gene="ABC-2 transporter Wzm"
 /note="putative"
 /pfam_match="PF01061; ABC2_membrane; 1"
 CDS
 18248..18847
 /blast_score=7e-61
 /blastp_match="AF285636.WZT Burkholderia mallei"
 /gene="ABC-2 transporter Wzt"
 /note="putative "
 /pfam_match="PF00005; ABC_tran; 1"
 CDS
 18952..20346
 /note="putative"
 /gene="glycosyltransferase"
 /blast_score=e-101
 /blastp_match="AF285636.WCBE Burkholderia mallei"
 /pfam_match="Glycos_transf_1"
 CDS
 20442..21167
 /note="putative"
 /gene="unknow"
 /blast_score=1e-26
 /blastp_match="AE009248.ATU3189 Agrobacterium tumefaciens"
 complement (21164..24301)
 /note="transporter domain"
 /gene="unknow"
 /blast_score=5e-29
 /blastp_match="AE009122.ATU1658 Agrobacterium tumefaciens"
 /prosite_match="PS00402; BPD_TRANSP_INN_MEMBR; UNKNOWN_1"
 complement (24351..27023)
 /note="none"
 /gene="unknow"
 /blast_score=5e-29
 /blastp_match="AP003581.ALR0267 Nostoc sp"
 complement (27806..29686)
 /note="none"
 /gene="cell surfarce protein"
 /blast_score=2e-34
 /blastp_match="AE010748.MA0851 2567 Methanosarcina
 acetivorans"
 /product="cell surface protein"
 complement (29535..30872)
 CDS

/note="none"
 /gene="cell surface protein"
 /blast_score=2e-36
 /blastp_match="AE010748.MA0851 Methanosarcina acetivorans"
 complement(30848..32647)
 /note="fragment"
 /gene="O-antigen"
 /blast_score=1e-72
 /blastp_match="AF105060.RFBC Riftia pachyptila"
 /product="O-antigen biosynthesis protein"
 /pfam_match="PF00535; Glycos_transf_2"
 complement(32574..35555)
 /note="putative"
 /gene="glycosyltransferase"
 /blast_score=7e-46
 /blastp_match="AE013462.MM2213 Methanosarcina mazei "
 complement(35533..36598)
 /blast_score=7e-37
 /blastp_match="AE013462.MM2213 Methanosarcina mazei"
 /gene="glycosyltransferase"
 /note="putative"
 complement(36516..37400)
 /blast_score=8e-22
 /blastp_match="AP003581.ALR0267 Nostoc sp"
 /gene="ALR0267"
 /note="putative ATP/GTP-binding protein, esterase fush
 9e-14"

Sequence 37507 BP; 5531 A; 13507 C; 13011 G; 5458 T; 0 other;

gatccacga	tggcgccag	cggattgcgc	agctcgtgcg	cgagcatcgc	gaggaaactcg	60
gtcacgtggc	ggccctcggc	ctccagctcc	tccatgcgct	tgcggtgcga	gaggtcgcgc	120
gtcaccttgg	cgaagcccat	gtgccttcca	ctcgcgtcgc	gaagcgccgt	cacgacgacg	180
tggcccca	agcgcgtgcc	gtccttgcgc	acgcgccagc	ccgtgtcctc	ggcgcggcgg	240
tgcagcagcg	catcgtgcag	ctcggcctcc	gggcggccgg	ccgcggccatc	ctcgggcgga	300
tagaagagcg	aaaagtgcgc	tccgatgatc	tcttcggcgc	tgtagccggt	gatgcgctcg	360
gcgccttgt	tccagctggg	cacgtggccc	gcgggggtcca	gcattagat	cgcgtagtcc	420
tgcacgccct	cgatcaggag	gcgcgagcgc	tctcgccttt	gccgcagctg	ctcctcgtgc	480
tgcgcggcct	cggtcaggtc	gcgcgtgacc	ttgcogaagc	cgaccagcgt	gccgtcctcg	540
tgcgcgacgc	ccgtgatgat	gacgttggcc	cagaagcgcc	ggccgtcctt	gcgcacgcgc	600
cagccctcgt	cctcgaagcg	gccgtcctcg	cgtgcgcgca	gcagctcctt	ggcgggccag	660
tgcgcgcgca	tgtcctcctc	ggtgtagaac	gcggagaaat	ggcggccgat	gatctcttcg	720
gcgcgatagc	ccttcaactg	ttcggcgccc	gcgttccacg	aacgcacgta	gccctcgggg	780
tgcagcacga	agatcgcgta	gtccgtgacg	gacccgacca	gcagccggaa	gatggcgtcc	840
tgcgcggcca	gcccgcctcg	ggcgccgaag	gcgaggttgc	gcaccgaggg	ccgcggaccg	900
gatgggcttg	gcatggctat	taggggtaca	cgggtgcacgc	aagtgtgcaat	cggacaatca	960
gagataattg	cttgatgaag	atgccatttt	tgggaattctt	acgcagggag	ctggggacgc	1020
tgcaggccca	gggcctgtac	aagtccgagc	gcgtgctcga	ctctccgcag	ggcgccagcg	1080
tgcaggtggc	cgcccacgac	gtgctcaact	tctgcgccaa	taactacctc	ggcctggcca	1140
accaccgcgc	cctggctcgc	gccgcgaagg	agacgctcga	gcgcgacggc	tacggcatgg	1200
cgtegggttcg	cttcattctgc	gggacgcacg	ccgtgcaccg	cgagctggag	acgcgcctcg	1260
cgcgcttctc	cggcaccgag	gacgcgatcc	tgtacggctc	ctgcttcgac	gccaacggcg	1320
ggctcttcga	gcgctgctg	ggcgaggagg	acgcggtgat	ctcggacgcg	ctgaaccacg	1380
cctcgatcat	gcacggcgtg	cgctctcga	aggcgaagcg	cttccgctac	gacaacaacg	1440
acatggcgag	cctcgaggcc	cagctgaagg	cggcggacgc	ggccggcgcg	cgctacaaga	1500
tgatcgccac	cgacggcgtg	ttctcgatgg	acggcatcgt	cgccaacctg	ggcgccatct	1560
gcgagctggc	cgggcgctac	cacgcgatgg	tgatggtgga	cgactcgac	gccgtgggct	1620

tcattgggccc	gcacggggcgg	ggaacgcccc	agcattgcgg	tgtcgaaggg	aaggtggaca	1680
tcctcaccgg	cacgctcggc	aaggccctgg	gcggcgcttc	cgggcggtac	acggcgggca	1740
agcgcgaggt	ggtggcctgg	cttcgcaacc	gctcgcgccc	ctatctcttc	tccaacacgc	1800
tgatgcccgc	catcgcgggc	gcgtcgtca	aggtgctcga	tctcctcgag	ggcgggcggc	1860
agctgcgcgc	gaagctcggc	cgcaacgccc	gccacttcgc	cgggcgagatg	acgcgcctcg	1920
gcttcacgct	ggcggcgccc	gaccatccga	tcatcccggg	gatgctgggc	gaggccccgc	1980
tcgcgaagga	gatggcgggc	cggctgctga	aggaaggcat	ctacgtgggc	ggcttctcgt	2040
ttcccggtgt	gccccgaggg	caggcgcgca	tccgcacgca	gatgtcggcc	gccccatgaac	2100
cgaagcacgt	cgatcgcgcc	atcgccgcct	ttgccaaagt	cgggcgcgat	ttgggagtca	2160
ttgcttgaga	accccttcga	agacgaagcg	cgagcccggc	atctggatgg	tcgagtcgcc	2220
caagcccgtg	gtcggccaca	acgacgtgct	gatccgcgtg	aagaagaccg	ccatctcgcg	2280
caccgacatg	cacatcttca	actgggacga	ctggctcgag	aagaccatcc	cggtgccgat	2340
gacggtcggc	cacgagtacg	taggcatggg	ggaggccatg	ggccaggagg	tgcggcgccct	2400
gcaggtcggc	cagcgcgctc	ccggcgaggg	ccacatcgct	tgcggccatt	gccgcaactg	2460
ccgcgcgggg	cgccgccacc	tgtgcgcgca	cacgcagggc	gtgggctgga	accgccccgg	2520
cgcgctcggc	gactacctgg	tgattccgcg	ggagaacgcg	ttcccgattc	ccgacgacat	2580
cccgacgag	atcgccctga	tcctcgatcc	gttcggcaac	gcggcgacac	ccgcgctctc	2640
gttcgacctg	gtgggcgagg	acgtgctgat	caccggcgcc	gggcccgatcg	gcatcatggc	2700
cgcgcccatc	gcgcgccacg	tcggcgcgcg	ccacgtcggt	atcacggaca	tgaacgacta	2760
ccgcctggcg	ctcgccacga	agatgggtgc	cagccgcgcg	gtgaacgtat	cgaaggaaaa	2820
cctgaaggac	gtgatgcgcg	agctcggcat	ggtcgagggc	ttcgacgtcg	gcatggagat	2880
gtcgggctg	ccctcggcct	tcggccagat	gctcgacacc	atgaaccacg	gcggcaagat	2940
cgccatgctc	ggcatccgcg	ccagcgaggc	cgcgatcgac	tggacccagg	tgatcttcaa	3000
ggcgctggcg	atcaaggcg	tcctacggcg	cgagatgttc	gagacctggt	acaagatgat	3060
ggccatgctg	cagagtggcc	tcgacctctc	gcccattggt	acgcaccgct	tcgacgtgcg	3120
cgagtacctg	aagggtctcg	agacgatggg	ctccgggaaa	tcggggaaag	tcgtcctgtc	3180
gtgggattag	cgaggggccac	ggatcgcggc	ccggttcccg	cggcggcgat	cgtgcaggcg	3240
ggttggcagc	acgccccggg	atttcgcgcg	tggtcacgt	gcgacttcgc	cgtcgacgca	3300
ccgatcgcca	tcgacgtcga	cgtgagcggt	tcgctgctcg	atgccagggg	cacctcgctg	3360
tggagcgagt	cgatggccgg	cagcgggcgc	gacacccgct	ggctaccgac	cggcacgtat	3420
cgcgctccc	tgcgccctga	gcccgttcgc	ttcccgcccg	cgcccgccct	ggtcgagttc	3480
gcgctctcga	tgccgggtgca	aggtgagcgc	cgcgctcggt	ccaccgcagc	gggctggtg	3540
cccgcgggcg	caccccacgg	cgtggcgcg	gcggcctggc	acctcgaggc	cctggatggc	3600
acgcccggcg	tcgagacgct	ggcctggtcc	gatcccacgc	acagctggtt	ctcccagcat	3660
ttcgaccatg	cgtcgcgcac	gtgcgtcgag	tacctggggc	ccggatggcc	cggttgggcg	3720
gggcgcggtg	tggatgcggg	atgcggcgac	ggcatcacgg	cgctcggcac	cgcgctgcgc	3780
tatgcgccc	agcaggtggt	gggcatcgat	ccggggcgct	gctaccgggt	gcttcccgac	3840
atcctcgagc	gcccacgccc	cgcgctgcgg	agcagcttcg	acaacctgca	gttccctccc	3900
gcggatgcca	acgcgctccc	gttcgcccgc	ccctcgctcg	acgtcatcgt	ctcgtgggtc	3960
gcggctcgagc	atttcgtggg	cggctacctg	gagctctact	ccgaggcgcg	gcgggtgctc	4020
aagcccggcg	gcctgctggt	catccatccc	ttccacctgg	acaccgcgca	ccacggccat	4080
cacctcgcg	agtacagccg	cgagcccttc	ctccacgacc	tgaagtgcgc	cgacgaggtg	4140
cgcgacatcg	tggtcgggcg	ggacgtgagc	ctcaatcgca	gcggcgggcat	cgcgccgacg	4200
cgcgcggagc	actggcgctg	gtacaacgag	tggcgccctg	tcacggccgc	gcagttcgag	4260
gacgagttgc	gcgcgctcga	cttcgagcgg	cgcttcacgg	cactgcgcgc	cgagccgctg	4320
gtggagtaca	agccggagct	gctgcgctac	tcagatgggt	agctgggctg	ctccgagctc	4380
tacgtggcct	gcatacaacc	caagcccgtg	gaattgaagg	cctacaggac	gcatgggcag	4440
gcggcattcc	ggtaagggtc	gccgggacgg	acaggagaac	cgcaagcgcg	ttggtatccc	4500
cgtcactccc	ggactgccgg	gatcaccctc	cttgccggga	ccccgatgaa	gacccagacc	4560
caggccgctg	tcgcccgggc	cctgctgtcc	ccgcccgtca	ccacgctggc	ccgcccga	4620
gccgtcacgc	tgacggggcg	gagcgaagtc	gtcaccgcga	cctcgagcgc	gaagggcagc	4680
ggcacgggtg	tcgtgaaggc	cgattgcacg	aaggccggcg	agatcacgct	caccggcatg	4740
accgcccagg	cggcccacat	ccacgaaggc	gaggcgagcc	ccaacggggg	cgtggcagtc	4800
ccgttcgtga	agaccgcgga	cacacggttc	acctacgtga	cgggcgccaa	gatgaccgag	4860
gcgcagtgcg	ccgcgtacaa	ggcggggcgg	ggcaagtagg	acgtgcacag	cgaggcccac	4920
aaggggcgcg	aggtccgcgc	ccagctggcc	ggccttgggt	gcagccgctg	caaaaagaaa	4980
aagcccgcgc	aagcgggctt	tttcggtgaa		ggggcgaaag	gattcgaacc	5040

ttcgaccccc	tgcaccccat	gcaggtgcgc	taccaggctg	cgctacgccc	cgaccgaact	5100
tgaaattata	cctgcatcgg	ggtcgcagcc	gccatgaatc	gaaacgcccc	gggcttcggg	5160
tagcatccgc	ccatggcgac	gacctcgtg	acgggtgggtg	caggctacat	cggcaccac	5220
atcctctgcg	cactggccca	ggccgggcga	cgcagcatct	gcacgacaa	ctactccaac	5280
agctcgccgc	gttccatcga	acgcgtgctg	cagatcgcac	cgggttgcgt	cgaggccttc	5340
gacgtggaca	tccgcgacgc	cgacggcatc	cgcaaggatga	tgcagggcg	cgacgtggac	5400
agcgtgatcc	acctcgcggg	actgaaggcg	gtcggcgaat	cggtcgagca	gcccagcgc	5460
taccacgaca	acaacgtccg	cggcaccgag	agcctgctcg	cggcgctcgc	cgactcgccg	5520
gtgcgcaagt	tcgtcttcag	ctcctcggcc	accgtgtacg	gcctcgccga	gaagatgccg	5580
atcgacgagg	acgcgcccac	ctcgcccgag	agcccgtacg	gccagaacaa	gctcgacatc	5640
gagcacatgc	tggtcgccct	cgccaagcgc	gacccctcct	ggcgcgtagt	gaacctgcgc	5700
tacttcaacc	cggtcggcgc	acacgagagc	gcgctgatcg	gcgaggacc	ggcgggggtt	5760
cccaacaacc	tcatgcccta	cgtgtgccag	gtgcctccg	ggcgactgaa	ggagctgagc	5820
gtctacggca	gcgactaccc	caccccgagc	ggcaccggca	tgcgcgattt	catccacgtc	5880
tgcgacctgg	ccgaaggcca	cgtggccgcg	ctggaagcac	tcgagcgcgc	ggccccggg	5940
accgtcctca	cgttgaatct	cgggacgggg	ctcggtaca	ccgtgctgca	attgctcgaa	6000
accttcgagc	gcgtgaacaa	actaagcatt	gcccggcgca	tcgtggggcg	gcggccgggc	6060
gacgtcgoga	tctgtacgc	gaacgcgggc	cgcgcgaagg	ccgtgctggg	ttggaccgcg	6120
aagcgcggaa	tcgaggagat	gtgccgcgat	gcctggcggt	ggcaggaag	gaatgcaaa	6180
ggctacgggt	gacggattgc	aaatcatggc	cttggcgacc	cttgctgcaa	tgcggcatga	6240
agggggcggt	tcttgaaga	tccaagggtc	gggatgcgtc	caggacgtca	aaaacagcac	6300
ccggcggttg	aaggaagaag	gcaaccctag	tgttgtcatg	gctccagacg	ggattgcgcg	6360
ccgagtaaga	atcgtcagtc	gagaccaact	cgctccggg	tgcaacgaat	cgtgagccga	6420
agcaacacct	cgcacgcgt	cgccgtcgtc	ggcattgcgt	tccgccttcc	cggcgctacg	6480
cggaatccc	tctggccctc	gctcctcgag	ggcgcaacc	tcgtcaccac	cgtcgatccc	6540
gcgcgctggg	cgaaggagcc	gtactaccac	ccgcgcaagt	ccgagcccgg	caccagctac	6600
tcgttcgccg	ccggcaccat	cggcgacatc	agcggcttcg	acgcgtcgtt	cttcggcatc	6660
tccccgcgcg	aggccgcgca	gatggatccg	cagcagccgc	tgctgctcga	gctctcctgg	6720
gaagccctcg	aggccgcggc	ggtgaagccc	tcgaccctgc	gcggcagcga	ttgcggcggtg	6780
ttcgtgggccc	gtccacccgt	cgactacagc	tacggcctcg	cctacgacct	ggcgctccatg	6840
gacgcgtcgt	cggccaccgg	caacaccgcg	agcatcgccg	ccaaccgcct	ctcgtatttc	6900
tacgacctcc	acggccccag	catggtcgtg	gatacggcgt	gctcgtcctc	gctgggtggcg	6960
ttccaccagg	cctgccagtc	gatcgccatc	ggcgagacgc	gccaggcgct	ggtgggcggc	7020
atcagcctgc	acctgcatcc	gtacgggttc	atcgcgttct	ccaaggcgte	gatgctctcg	7080
aagcgcggcg	cctgccgcgt	gttcgacgcc	agcggcgatg	cgtaagtcgc	ctcggaaggc	7140
gcggggcggtc	tgggtgctgaa	ggacctcgac	gcgggcgct	ccgacggcaa	cccgatcctc	7200
gccgtcgtcg	ccgcgagcgg	cgtgaatacc	gatgtgtacg	cctccagcct	caccgtgccc	7260
agcgtcgatg	cccaggtgc	attgctgtcg	accggcaccg	agcgcgcggg	catcgacccg	7320
tccgagatcg	actacctcga	ggcccacggc	ctgcgcgcgc	tgggtgggcga	tccgatcgag	7380
acccgcgccc	tcggcatcgc	gctcgggcgc	gagggcgcc	gccaccagcc	gctgccgatc	7440
ggttcggtga	agagcaacgt	cggccacctc	gagggcgcc	cgggcatggc	gggcctggtg	7500
aaggcgctgc	attgcctcaa	gcaccgcgcg	gtgcgcgcca	ccatccacct	cgattcgccc	7560
aaccccaaca	tccacttcga	cgagtggaa	ctgaagccgg	tcctcgagcc	gctcacgctc	7620
catccgtaca	agaagctcgt	cgtcggcggtg	aactcgttcg	gcttcggcgg	cgccaacgcg	7680
cacgtgatcc	tcgagagcac	gcccacttcg	cccaaggcgg	tcgaggcccc	ggcggcgag	7740
tggtcgctca	tgctgagcgg	ccacgacgac	gcggcgctgc	gcgacgtggc	ggcctcgta	7800
gcgaacttcc	tcgagggcag	cgagcaaccc	gcgctgcacg	acatcgcgta	caccaccgtg	7860
cactgccgcg	actggcacca	gctccgcgcc	atgggtggtcg	gcaccgaccg	cgccagcatc	7920
gccggcgccc	tcgcgcaatt	cgccaagggc	cacaacacgc	ccggcgctcgt	gagcggcgt	7980
gcgtgccc	acgcgtccgg	cccgccttcg	gtgtattcgg	gcaacggcgc	gcaatggccg	8040
gcaatggggc	gcgcctgct	cgcggcgag	ccggtgttcc	gtgcgcgcgt	gcaatcggtc	8100
gacgcgtgt	tccgcccgt	cggcggttac	tccatcctcg	agatgctgga	gcaggaggtc	8160
cccgcgacc	agctggaccg	caccgagatc	gcccagccgc	tgctcttcgc	cgtgcaggtg	8220
ggcttgaccg	agctgctgcg	ccactggggc	atccagccca	cgcgcgctcac	cggccacagc	8280
gtgggcgagg	tcgcccgcgc	ttgggcgagc	ggcgcgctga	cgctcgacga	tgcggcgcg	8340
gtgatctacc	accgcagccg	ctaccagggc	gagaccaagg	gcgcgggcgc	gatgaccgcc	8400
gtcggcctcg	ccgaggacgc	cgcgcgcgcg	cagatcgagg	cgctcgggtt	gaaggcgag	8460

gtgaacgtcg	cgtgcatcaa	cagcacgcgc	aacgtgacgc	tcgccggcac	gcgcgccggc	8520
atcgaggcgc	tcgaggccga	gctcacgcag	cgcaaggtgt	tccaccgcgc	cctcgacctc	8580
gactacgcgt	tccacagccc	ggcgatggac	ccggtgcgcg	acgggctcgt	gagcgcgctg	8640
cgccggcctca	cgccgcgcgc	cacccggttc	gcgttccatt	cggcggtcac	cggtcgcgcc	8700
gccgcgggca	accagctcga	cgccacctac	tggtaggagca	acatccgcga	gccggtgcgc	8760
ttccaggccg	cgatccgcgg	catcgccgag	tcgggctca	atgtcttcat	cgagatcggc	8820
ccgcacccga	tcctcaagaa	ctacatcaac	gacggcctgc	gcgcgccttc	gatcgagggc	8880
cgcgcgctgg	tcacgctgca	gcgctcggcg	agcgatcgcg	cggcgatccg	cgccgcgcgc	8940
caggaagtcc	tgatcacccg	ctgcccgggtg	gagaccgcga	agatgctccc	cgcgccaggc	9000
cacttcgctcg	agctgcgcgc	ctaccgctgg	cagcgcgagc	gccactggcg	cgcgccacc	9060
tcgcaggcct	acgacctcat	ccagcacggc	aagcagcacc	cgctgctggg	ctatcgctcg	9120
cacgagaacg	acttccagtg	ggaaaaccac	atcgacaccg	cgctctatcc	ggcgtagcgc	9180
gaccatgtcg	tcggcgccgc	cgtgggtgttc	cccgcggccg	gcttcgctga	gatggcgctc	9240
gccgcctcgg	cgatcgcgct	ggcgccgcga	gcgcacgaga	tcgagacgct	cgagatccgc	9300
agcgcgctga	tgctcgagga	ctccacgtcg	aagaccgtgc	gcttcgcgct	ggaaccgat	9360
ggcgcttcca	tgatccgcag	ccgcgctcgc	ctgagcgagg	atccgtggca	actccacgtg	9420
gtcggcaagc	tggtcgccac	gcccaccgag	cttcgcgcgc	cgccgcgcga	gtcggcgccg	9480
tcgcgcaagc	ccgatgtgat	cgcgcccag	cactacgaag	gcgcgcgcga	ggccggcctc	9540
gcctacggcc	cgcccttcca	gtcgggtctcc	aaggtgtggc	tcgacgccgc	ggaccggggc	9600
agcgccctacg	cgcgccctgat	cctccccaa	ccgattcgcg	gcgagctggg	cgtgatgcac	9660
ctgcaccccg	cgctcgctgga	cggtgcttc	cagctgctgg	tcgacctgct	gcgcgcgcag	9720
gcctcccggc	acgcgcagat	cgccctcgtg	ccgatccctg	tcggggcgac	gcgcctctac	9780
ggccctgcgc	ccctcgtcac	cgccctgcgc	gtccgcctga	ccgcgcgcag	cccgcgctcg	9840
ctggctcgcg	acttcgagct	ctacggcccc	aacggcgacg	tggtggccgc	gctcacccggc	9900
gtgcgcttcc	gcggcgctcca	gctgaagcag	ccgcaggggc	cgcgccctgc	ctacctgtcg	9960
taccgcgcga	tcgcgcgcct	gaatggcggc	gacacgcccg	gcgcgaagcc	gctgcccctc	10020
gcgttgctgg	ccgcggcgctg	cggcgagcgc	tggcattcgc	tctcgcccg	ctcgcgcaag	10080
cgctactacg	atgaggctga	gccgctctcc	gacgtgctgt	gcagctcctt	cgccgagcac	10140
gcgctgcgaa	agcggtatcg	cgaggccccg	ctcatcgacc	cgccgcgcgt	cgccgcgctcg	10200
acgcgcggcc	agcattggcc	gctcctcaag	cgcgcgatcc	agatgctggg	cgaggaccag	10260
ttggtcgcg	ccgcgcgaag	cggtggcggc	tggagccccg	tcaccgaatt	gcccgcgcgc	10320
caggagagct	ggatcacgct	gctgcgcgac	taccgggacg	aggccggcca	gctgatcgcg	10380
ctcgggcgcg	ccggcgctca	cctcgcgga	gtgttcgcgc	gcaagggaagc	gtccttgccg	10440
agcctgttgc	cgccacgcga	ggcgaacgtg	ttcgcgcgcc	tgtgctcggg	cgccaccgcg	10500
ttcgccaaacg	ccgggcgcgc	catcgccgac	acgctggcg	tcgcctcga	gcgcgctccc	10560
gcgcacccggc	cgctgcgcgt	gctcgaggtg	acgcccgtgc	gctccgagat	cacgctgcgc	10620
gtcctcgagg	ccgtggacct	cgaccgctgc	gagtagccag	tcggctgcac	cacggacgaa	10680
gcactgggag	agtagcgaag	cgctcctcgat	gccattccgc	acgtcgagac	ttgcgtcgctg	10740
gacctgaaga	tcgccggcat	gggcctgaag	gcgcgaacc	agggcccggt	cgatgtcgctg	10800
atcgtttccg	atggcctgct	cggtgcggcc	gatccggatg	cgccgcgcggc	gcacctcgcc	10860
ggcgcgctcg	cggaaagacg	gctgctgggtc	atggtcgcg	agcatccgtc	gcgctggacc	10920
gacctcctct	tcggcgcgga	cccggcggtg	tggacgctcg	ggccctcggg	ctcgacgcgc	10980
tcgcgcctgc	gttcgcccga	cgagtggcgc	atcgcgctgg	cccacctcgg	cttcggggccc	11040
accgtcgcg	tgcccagagct	gcccggcctc	gagcgcggtc	cgtagctcct	catggcgcg	11100
cgcgacggcc	cggtgcgcga	ggctgcccag	cgcagcggtc	ccgcgaagtc	gggctggatc	11160
ttcggtgcag	aagcggggcg	ttattcgggc	gcgttcggcg	cctgcgtgca	ggcggtatctg	11220
cgcgcacgcg	gccacgcggg	gttcgagatg	gcgcccagatt	ccgtcgaggg	ggccgagatg	11280
gcgcagcttg	cgcgcgcgcg	ccagacgctg	ggcgacatca	gcggcatcgc	cttcctggcc	11340
ggcctgccc	acgacgcggc	cgccacgggc	gacgcggccg	agcgccctcg	gcgcagacc	11400
gcgcgctgcg	ccgcgctcgc	gcgcctgctg	cgcgaatgca	aggcgagtg	catgacacc	11460
gcggtgtgga	tcgtgacctc	gggcgcggcc	tcaggcctgg	cgcgcaacgt	ccccgccctt	11520
gcctcgcgcc	gcgcgccctc	cgatccggac	gaggcgatcg	tatgggggtt	cgcgcgaccc	11580
gcgatgaacg	aattccccga	gctgcgcctg	cgccctcggtg	actgcccga	cccggagaac	11640
ctcgagcgca	acgcgcctc	gctgggtgcag	gagctgctct	cgccacccgc	cgacgacgag	11700
gtggtcctca	ccggcgccgg	ccgctacgcg	ctgcgcctgg	gcgcgggtg	gcccgcgcga	11760
tcggtgcgca	ccgctaccca	gggcatcacg	cgccctcgact	tcgcgcgcgc	cgcccgctg	11820
aagaacctcg	cgtggcgcg	cgacgtgcgc	cgcaagccgc	gcgcgaacga	ggtggaaatc	11880

gaggtgcgcg	ccgcgggcct	caacttccgc	gacgtgatgt	acgcgatggg	cctgctctcc	11940
gacgaggcgg	tcgaggcg	cttcgccggc	gcctcgctcg	gcatggagct	ctcgggcgctc	12000
gtcaccgcgg	tgggcaagga	cgtgacctcg	gtcgcgcgcg	gcgacgaggt	gctcgcgcttc	12060
gcgccctcgg	cgttcgccac	gcacgtggtc	accaccgccc	attcggtcgc	gaagaagccc	12120
gccggctgga	cgttcgagtc	cgcggccacc	atcccgaccg	cgttcttcac	cgtgtactac	12180
tcgttgaagc	acctggccca	gctgcgcgag	ggcgagcgcg	tcctgatcca	cggcgcccgcg	12240
ggcgggcggtg	gcacgcggc	gatccaggtc	gccaaagtggc	tgggcgcgga	gatcttcgcc	12300
accgcgggct	ccgacgagaa	gcgcgacttc	gtgcgctgc	tgggcgcgga	ccacgtgctc	12360
gactcgcgca	cgtcacggtt	cgcgcacgac	gtgctgcgca	tcaccaacgg	cgaggcgctc	12420
cgctgggtgc	tgaactcgct	ctccggcgag	gccattgcgc	gcaacctgcg	cgcgctgcgc	12480
ccgttcggcc	gcttcacgca	gctcggcaag	cgcgactact	acgagaacac	gcacatcggc	12540
ctgcgcccgt	tcgcaacaa	catctcgta	ttcgggcgctg	atgccgacca	gctcatgaag	12600
gagcgggccc	acctcgcgcg	ccagctcttc	aacgagctga	tcagactctt	cgagcagggc	12660
gtgctctcgc	cgtgcctta	ccgcgccttc	gccgccaccg	aggcggtgga	agccttcgcg	12720
tacatgcagc	actcgcgcca	gatcggaag	gtgggtgctgt	cgttcgccga	cggcgtgaag	12780
gcgcagcccg	cccgcccg	cgtgcgcgc	gagctggcg	tgtcgccgaa	cgcacgtac	12840
ctcgtcaccg	cgggctgtc	gggcttcggc	ctgcgcaccg	cgagtggtc	ggtggacaag	12900
ggtgcggcc	acctgggtgct	ggtctccaag	agcgcgcccg	agtcgctcga	gaccaaggcc	12960
gcgggtggccg	acctcgaggc	ccaggcgccc	acgggtgatcg	cggaggcgctg	cgacatcacc	13020
gaccgcgcct	ccgtgcaacg	gctgctcgcc	gaggtcgcg	cgcgctgcc	gccgctgcgc	13080
ggcgtgatcc	acgcggcg	cgtgatccag	gacggcttca	tcaccaacat	gacggccgcg	13140
cagatccgg	acgtgctggc	gcccgaagtc	ctcggcgc	gccacctgga	cgagctcacg	13200
cgcgggcgca	agctcgattt	cttcgtgctg	tactcgctcg	cgacgacgct	gttcggcaac	13260
ccggggcgca	ccaactacat	cgcggccaac	tgttctctcg	aggcactcgc	caaggcgcg	13320
cgcgtcgagg	gccagccggc	gctgtgctg	ggctggggcg	cgatcgggga	cgtgggctac	13380
ctcgcgcgcc	acgagaaggt	gaagggaagc	ctgcagacgc	acatggggcg	caccgcactc	13440
gagtcggagg	ccgcgctcgc	ggtgctggag	cagctcctgc	tcgccgacgc	ctcgggcctc	13500
ggcgtgctcg	acttcagctg	gcgcaacctc	cgcgcgttcc	tccccaccgc	cgcctcgccg	13560
cgttcccg	agctgggcgc	gcgcggcg	gacgcgaagg	acgacgacga	ccggcgcgcc	13620
gagctgcgcc	gcctcgccaa	cgagctggac	gcctccgagc	tctcggcgat	gttcaccgac	13680
ctgctgcgc	gcgaggtggg	cgagatcctg	cgcacccgc	cggaaccgct	cgatacgcg	13740
cagcccttgc	aggaaatggg	catggattcg	ctgatggggcg	tggagctgct	gaccgcgggtg	13800
gaggcgcgct	tcggcggtga	cctgcgggtg	atgtcgctct	ccgagcagcc	gtcgatcgag	13860
aagctggctg	accgcacgtg	gcgcgcgctg	aaggaccgca	acgcccgcac	cgaggccgaa	13920
tcgcgcagcg	accagatcga	gcgcgtcgcc	gcgcagtaag	cgcccagact	cgactcgcg	13980
caggtcgagg	agctcacgca	ggcggtcgat	gacgcgcaag	ccacgcagca	gggacgtccg	14040
tgaaggcgcg	caacctccgc	ggcctctcgc	aggcagtgaa	ggagcagctg	atccagcgcg	14100
tgctcgagca	gcgcgtgcgc	cgggctcgaga	aggacaagcc	ctcggcgggtc	gagtcgcgca	14160
tcaacgcctt	cgcggcgac	cccgatgccc	tgggccaacat	ctcgagcg	tactgcccgt	14220
tcgacatgca	cccggtctac	cagcagatgc	agctggtgaa	ggaaggtgcc	gcgaagctgg	14280
gcattcccga	tcctgacttc	cgcgcgcacg	acggcggtggc	cgcggatacc	acgcgcattg	14340
ccggggcgga	gtacatcaac	ttctccagct	acaactacct	cggcttctgc	ggcgacccgg	14400
ccgtggatgc	cgcggccaag	caggccatcg	acgaatacgg	gaccctcggt	ctcgggcagc	14460
aggctcggtg	ccggcgagcg	gccgctgcac	cgcgagctgg	agcgcgccat	cgcggaggctc	14520
tatggcggtg	acgacgcccgt	ggccttcac	agcgggccacg	cgaccaacgt	ctccaccatc	14580
ggccacctgt	tcggggcccg	cgacctcatc	gtgcacgacg	cgttcgtgca	caacagcatc	14640
ctgatgggca	tcagctctc	gggcgccaag	cgcagtgcc	tcccgcacaa	cgactggcg	14700
gccctggacg	agctgctgcg	cgcgcagcgc	cgccacttcg	agcgcggtg	gatcgatgatc	14760
gagggcgctct	acagcatgga	cggggactac	cccgagctgc	ccgaattctc	gcgcctgcgc	14820
cgcggccacc	gcgcgttctt	gatgggtgat	gaagcgact	ccttcggcg	gatgggccc	14880
cgcggcttcg	gcacccgcga	ccacttcggc	atggccggcg	acgaggccga	catctggatg	14940
ggcacgctct	ccaagacgct	cgcctcgtgc	ggcggttca	tcgccggcg	gaaggcgctg	15000
gtcgagcacc	tgaagtctgc	cgcggcgccg	ttctctaca	gcgtcggcac	ggccccgc	15060
gtggcgggcg	cgcgcgtcgc	cgcgcgtgaac	cgcagtcgcg	aggagccggc	gcgcgtggcc	15120
gccctgcagg	cgcgcggacg	cttcttcttc	gagggccgcg	gcgcggcg	catcgacgtg	15180
ggcctctcgc	agggcatggc	cgtgggtgccc	gccatcaccg	gcagctcgat	ccgcggcg	15240
cgcctggccg	aggcgatggt	ccagcgcg	atcaacgtcc	agccgatcgt	ctaccggcc	15300

gtgcaggaga	actcggcgcg	cctgcgcttc	ttcgtgagcg	ccacgcacac	cgaggagcag	15360
ctgcgcttca	cgggtccgcga	gctcgcgag	gcctggcgca	agctctgagt	ggcggggccc	15420
cggctgcgcg	tgcgtcgtctg	cacccacggc	ggcctgaacg	gcgcgctggt	gctctcgcgg	15480
ctgctggccg	ccccacgct	cgaagtctcc	gcgctggtga	tctccagccg	cgcgcgcggt	15540
gcccattgagt	cgtatggccg	tgcgcgcgtc	ggatacgtgc	gcgcgagcgg	cgtcgcctat	15600
gcgctgtacc	tctgggtgcgc	cacggcgctc	gcccacgtgc	tgctgcgcgg	cacgtccgag	15660
ggcccggtcg	cgcgcacgcg	cctcgcgcgc	ggcatcccgc	tccctgccac	gcccgcgctg	15720
aacgatgccca	ccgcgcgcgc	cttcacgcgc	ggggcggcgc	cggacctgat	cgtctccgcg	15780
ttcttcaacc	agcacatcga	cgcgcacgtg	gctgcgctcg	cgcgcgtggc	cgcggtcaac	15840
atccaccctg	cgcgcgtgcc	gcacttccgc	ggcgtggatc	cgggtgcctt	cgcgcgcctg	15900
cgcggtgccg	agcgcacagg	cgtgagcgtg	catcgcacgc	aacccggctt	cgataccggc	15960
gcgctgctcg	cccaggaaac	cgcgcgcgag	gcccccgcca	gcgtctttgc	cgccaccgcg	16020
gcgctctatg	accgcggggc	ggcgtcctc	gcccggcgtg	ccgcggcgct	ggccgcgcgac	16080
ccgcgcggaa	ccccccagcc	ggcggggggc	tcctacgact	cctggcccac	ccgcgcggcag	16140
gtcgcgcgct	tcgcgcgcgc	cggcggggcg	ctgctccgcg	cccgcgacct	gtggcgccctc	16200
gcgcgcgggg	ggcggggcg	tttcgtaata	gaatcagcgc	ggtagcggcg	actcacgcgc	16260
cgcttttccc	cgcgcgcgcg	aggcccatga	aacactggct	gaagcaacac	gcaatcttcg	16320
cgctcacctg	cctgctgcgc	accgtggcgc	cgcgcctcta	cttcggcctg	atcgccctcg	16380
acgtctacat	ctccgaatca	cgccttcgtg	tgaggagccc	ccagcgccag	gtgcagaccg	16440
gcctgggtggg	cgcctctgct	tcgggcacgc	gcttctcgcg	ctcccaggac	gacacctact	16500
cgggtgcacga	cttcacacac	tcgcgcgacg	cgcgtggcga	gctggacaag	aagctcgcgc	16560
tcgcgaagct	ctacacggcc	gccaacatcg	acttcaccaa	ccgcttcccg	gggctcgact	16620
gggacgacag	cttcgagggc	ttccaccgct	actaccagaa	gaaggtcacg	atcgacttcg	16680
acaccgcgct	ctcgatcacg	gtgctgcgcg	tcgcgcgcct	cgagaaggcc	gactcgcggc	16740
gcatcaacga	cctgctgctg	cagatgggcg	agcgcctggt	gaacgagctg	aacgagcgca	16800
gccgcacagg	cctgatccgc	ttcgcgcagg	ccgaggtctc	gctcgcgcgag	gacaagggtga	16860
aggatgccgc	gctggcgctc	tcgcgccttc	gcagcaacca	gtcgggtgtc	gagcccagac	16920
gccaggcctc	gatccagctg	caggggcggt	ccaagctcca	ggaggagctc	atcgccaccg	16980
aggggagcgt	ggcacagctg	cgcgaagctc	cgcgcagatc	ggcgcgctcg	gtgaccggcg	17040
agaacaagtc	ggcgcgctg	cgcgtggcga	tgccgcgcga	gtccgccaag	gtgaccggcg	17100
gcagcggtc	gttcagcgcg	cgcgcgcccg	cgttcgagcg	cctcacccctc	gagaagggtc	17160
tcgcgcagcg	ccagctgggc	gttgccctca	ccgcgcgtga	gaccgcgcgc	agcgaggcgc	17220
agaggaagca	gctctacctc	gagcgcatcg	tcgagcccaa	cctgcccagc	gaatcggtcg	17280
agccgcgcgc	catccgctcg	atcttcaccc	tggtcgtgct	gggcctggtg	gcctggggcg	17340
tggtgagcct	gctgggtggc	agcgtgcgcg	agcacgtcga	ctaggccgtg	gagcagaacc	17400
cttccctcct	gcgcgcctc	gggggtgcag	ggcgcgtgct	ctagcgctcg	ctgatggcgc	17460
aggtcatcac	gcgcttcggg	cgcgacgacc	tcgcgcgtgct	gtggctgggtg	gtcgagccga	17520
tgatcttcac	gctcggcggtg	acggccctgt	ggaccgcggc	cggcatgaac	cacggctcgt	17580
cgcgtgcgat	cgtggccttc	gcgggtcacgc	gatactccgc	ggtgctggta	tgccgcaact	17640
gcgcctcgcg	cgcctcgatg	gccatcgagg	ccaacaaggg	cctgctcttc	caccgcaacg	17700
tcgcgcgtgat	cgcgcgtgtc	gtgacgcgca	tcacccctga	gatcgcgggc	gccacggcgt	17760
cgttcaccgt	gctggggcatc	ttcttcacat	ggatcggtcg	gatgccgatg	ccgatcgaca	17820
tgctcaagggt	cgcgttcggc	tggttcacat	tgctcgtggt	cggcgccagc	ctggcgctcg	17880
ccatcggcgc	cggcacccgc	tactcgctgg	tggtcgagcg	gctgtggccg	ccgacggcct	17940
acctgctggt	cccgctctcg	ggcgccgcct	tcacgttcga	gtggctgccc	gagaaattcc	18000
agcagttcgt	gctcttgctg	ccgatgggtc	acggcaccca	gctgctgcgc	gacggctact	18060
tcggcaacgc	ggtgcgcacg	cactacgacg	tggttcacat	ggcaacggtc	tcgctggggc	18120
tgctcgtgct	cggcatgcac	ctggtgcgca	tggtcgccaa	gcgcgtggag	gccgcgtgat	18180
ccgggtcgag	aacgtgacca	gatgtactca	cgcgcctcgg	cccagaacgt	gctgacgcgt	18240
cagcttcgag	ctcggccgcg	ggcgcaacat	cggcatcctc	ggcaagaacg	gcgcgggcaa	18300
gtccacgctg	atccgcctga	tcagcggcgc	cgcgcgcctc	acaccggggc	gcacgtgcgc	18360
cgagatgagc	gtgtcctggc	cgtcgccttc	cgcgcgcgcg	ttccagacct	acctcacggg	18420
cttggaacaac	ctgaagttcg	tctgccgcac	ctacggcgctc	gactaccgcg	acaaggctcg	18480
cttcgtcgag	gacttcacgc	agctcggcgt	gtatttcgcg	gagccgggtg	tcactacttc	18540
gcacggcatg	atcacgcgcc	tcgccttcgc	gctgtcgatg	gcgggtggag	tcgactgctt	18600
cctgatcgac	gaggcgatgg	tggtgggcga	cgcgcgcctc	cacgagcgct	gccacgtcga	18660
gctcttccac	aagcgcaagg	accgcgcctt	catcctcgtc	acccacgacg	ccaagggtgat	18720

caagctctat	tgcgagagcg	cctgcgtgct	gcacgagggg	cggctgctgc	ccttccccac	18780
cgtggacgcc	gcctacgagt	tctacatgaa	cgaggtcatg	caggacctcg	ccccggaggt	18840
tgctgaacg	cccccgcgcg	gccgctgcmc	gtgctcctcg	agctgcgccg	ggcgttcgac	18900
ggccacgccc	gcatcccga	ggaggcacgc	ctgctgttcc	gcgccctgcg	catgatcgag	18960
ggcatcgacg	tggagggcct	gctgcagcac	agcgcccaag	tgctcgccaa	ggcctgccc	19020
ccgcgcggcg	gcggcgacct	cgcaccggac	cgccagctga	accgcctgtc	gcgcgtcgtg	19080
gtctccacca	agcaggagct	caccaacgcg	cacgtggcga	cggcgctgat	ggccattcgc	19140
cagctgctgg	gcgcgcgcga	gaccctctac	cgcttcgacg	cggcgcaactt	ccgcgacttc	19200
gtctggcagg	cgctcttcgc	gcgcacgctg	cacgcgcggg	acttcgactc	ggtcacccgc	19260
gccggctttc	gcgtcgcgcg	cgtgccgtgg	acgggcatgc	accgctgcgc	gctgggtcacg	19320
cgcaagctcg	gctacacgct	gcatccgcgc	atcgacactt	ccgatttcga	cgtgctgac	19380
gccgagacgc	cctaccgggc	gcgggctcgcg	gacggcacgc	gcctcgtggg	gcgctaccac	19440
gacgcgatcc	cgctcctcat	gcccgcacacc	atctccgaca	tgctcctacca	ccaggcctcg	19500
cactaccggg	cgctgcgcgc	caacgtggcc	tcgggcgcgc	acttcgcgtg	cgtctccgaa	19560
gccacgcgca	aggacctgct	ctcggctctt	cccgaggtcg	aggcgcgctc	cagcaccatc	19620
cacaacatgg	tctcgaccca	ctacttcgcg	gagaccgccc	acgcccggcg	catcgaggag	19680
atcctgcgaa	cgcgcgcgag	cgagcgcac	aaggcgcgcg	ccaacggagc	cgccaccgcc	19740
gtcgacgat	cgccctcctt	cctcgcgcgc	gccgcgcaac	ccagcgaccc	cggctacctg	19800
ctcgccgtgg	cgacgatcga	gccgcgcaag	aaccacgcgg	gactgatcgc	cgcgtgggag	19860
cacctgcgca	cctcgcgcctt	cccgggcctg	cgctgggtgg	tggtcggcat	gcccggctgg	19920
ggctacgagg	cgctggtagc	caagttcaag	ccgtggctcg	cgcgccggca	gctcttcgtg	19980
ctcgaggacg	tgcccgcgcc	cgagctgcgc	ctgctctact	cgcacgcgcg	tgccacggtg	20040
tgcccagact	tcggcgaggg	cttcgatttc	tcgggcgtgg	aggcgatgcg	ctgcggcagc	20100
cgggtgatcg	cctcgggat	cgccgcgcac	cgcgaggtgt	accgcgacgc	cgccgagtag	20160
tgacgcccct	attccgtggc	cgacctcgcc	gaggcgatcg	ggcgcgctcat	cgaccgcggc	20220
gcgacggggc	tgcgccaggc	gctggtcacg	cgcgccaccg	aggctctcgca	gcggtacacg	20280
cccgaagcca	tcttgccgca	gtggcgcgag	tacctgctcg	gcacgggtgcg	cgcgagggcc	20340
ccgtgagcga	tccgaaggcc	accgaccagg	cgtggggagga	gtggggccag	cgcgaccctt	20400
acttcggcgt	gatcaccaac	ccgcgccttc	gccgcgggca	gatggacgag	gaagcgaagc	20460
gcgaattcct	cgcgctccggc	cgctccatg	ccgactacgt	gatgcgcacg	gtgcacgcgc	20520
acatcgcgcc	cgcttccgt	ccgcgcacca	tcctcgattt	cggtgcggc	gtgggcccgc	20580
tcgtcatccc	cttcgcccgc	caggccgagc	aggtagaccg	cgccgacgtc	tccccctcga	20640
tgctggccga	ggctgcgcgc	aactgcgcgc	agcaggggcgt	ggccaatgcg	cgctcgtgg	20700
tctcggacga	ctcgctcacc	gggctgccc	gccccttcga	cctcgtgcac	tcgttcactg	20760
tggtccagca	catcgatccg	gcgcgcgggc	gcgacatctt	ccgcgcctg	ctgggacacc	20820
tggcgagcgg	cggcgtgggc	gcgttgact	tcgtctatgc	gaagcgtatc	tacggggcga	20880
cgtacggcgt	ggcgccaccg	cccagccgcg	gccgcggccc	gccgcggccg	cagcctccac	20940
ccagccgggc	ggagatcaag	gccgcgagca	aggcgcgggc	cgcgctggcc	acgcaggcac	21000
ggcgccggga	accggcacc	aatcccagacc	ccgacatgca	gatgaattcc	taccggcgcg	21060
cggagatgct	gttccctcgtg	caggaagccg	gcgtcacgcg	cttcacagtc	gagttcacccg	21120
accacggcgg	cgagctgggt	ctcttcctct	tcttcgcaa	gccctaggcg	tccttcgagg	21180
cgaaggagag	gcccagcgcg	gtgatcgtcg	cggcgtgctg	cggggcgacc	gcccgcgcgg	21240
ccgcgacttc	ggcagccacg	cgctcgaggt	cgccgcgcga	cccgtacagg	cgcacgcgcc	21300
agtagtgac	ccagtaggag	tccggacagg	ccgacgcgat	tcccagcgcg	cgctcgacct	21360
cgggcccgcg	gcgcccctgg	tgcaactggg	gcgaggccca	ttgcagggtc	acggaggggct	21420
cgagggggcag	gacggcgagc	agcgccctcg	tcaccggccag	cgcgcgcgcc	tcacgcgcgt	21480
cgctggcgag	gcgcacacgc	gtgggcaggg	actccggcgc	aagcgcgctg	gcggccgcga	21540
gcgaatcgcg	cagcatcgca	gcgtcgccgc	gcgcgagcgc	caggcgggcg	cgggtggtagt	21600
gggtccagaa	cgaaggcgcc	gcggcatcga	gcgcgcgcgc	gacctcggcg	cttcgctcgc	21660
ccccggagag	ccgggacgcg	gcccacaaca	ggttcgcctc	cgcgcgctgc	gcccggcacgc	21720
tccgcaacgg	cgccagcacc	gcgggtggcga	cctcgcgcga	accctcgcgc	gacgagaagg	21780
ccgagcgcg	gaactgcac	agcgattcgc	cgagcgacga	atgttcggcg	cagagccgcg	21840
ggggatccac	cgcgcggagc	gcgcggggcg	cacgcagcgc	cgctcctcgc	tcaccgcgaa	21900
gcgcgtgcag	ccgcaggcga	tggaaacgcg	cccactccgc	cgacatcgac	gcccggccgc	21960
gctccagcgc	accctcgacg	gcgccatcgg	tgcccccgcg	cgcgatgtgc	aggatggccc	22020
aggtcgcctg	cgcgccgggg	aggttcgtgc	cttgcgcgac	gagcggggtcg	aggaccgcga	22080
gcgcgcggcg	ggcgtcgcg	gcgccactga	agtcgttggc	caggccgatc	agcgccccgg	22140

cgagctcggg	ggccgcgcgg	aaggcggccg	ccgcgtcgcg	ccccaccgcc	gcgaggtcgc	22200
cgcgagccg	gtagaggcgc	gcgcggtggt	aatagcccca	gaacgggtcg	gggctcgcga	22260
tcagtgaatc	gaacgagcgc	tcgacctcgc	cgcccacgtt	gccgaggtcc	acctgggtcgc	22320
cgccccacag	ggtcgccgca	tcgcgcggcca	tcggctcgtc	ggggacgaag	gcgcggtgga	22380
tcagctcgac	gtcgccgaag	cgccgctcgc	gcaacgcgag	ctccagggcg	ttcagccgct	22440
cgaggctcgc	cttgtagctcg	cgataggcat	cgcggtcgcg	gtcgtattcc	tcgcgcggcac	22500
tgtcgtccat	cagcgaggcc	gtgagccgct	ggtcgccgtg	aacgggtgtcg	tatgcgtagg	22560
ccgcgcgatc	gtcgtagggc	agcgtggccg	cgcgacgcga	cccgaacgcg	gcgcggtcgc	22620
ggaatcccag	cttgcccagc	acgcgggggt	ccaggctcgt	ggcttccatc	tcgcccgggc	22680
tccagtcgcg	cagcgccggc	acgttcgagt	agagccacaa	cggcgtgcgg	tagagctgcg	22740
ccaggtcgcg	gtccggggcg	ccctgcgcct	cgaagatcgc	cttccactcc	gcgagtcctt	22800
cgtggccggc	ggagaagatc	ggcacgtggt	cgccgaagat	cgccagcacg	tagggccgcc	22860
cgcgcacctc	gagacgctcg	accagcgagc	gcagcgcgcg	gtcggtttcg	gccatcaccg	22920
cgcagtaact	gtcgatcgcg	ccctggcgca	cgtgcccggc	gaggaagtgc	ggcgccacgt	22980
agcgcggctc	ggcgcccttg	tcgcccaggt	agggctcgtg	gccccatcag	gagacgcgca	23040
acacgaagcg	cgggccgcga	tcgtcgagct	ccgcgatcac	ggcatccatc	acggccgaat	23100
ccgcgatcg	ctcgaactgg	cgcggtcgat	gggcgaacgc	ctcgtcggcg	tggaacgcgg	23160
cgaagccgag	cttcttgtag	acgcggtcgc	ggttgtagaa	ccagcggtgg	aacggatgga	23220
tgccacgct	gcggtagccg	tgggcgcgca	gcgtgcgcgg	caccgcatac	acccggtcgc	23280
gcaactgccc	catgtagggc	agcgcgcggc	agccttcctc	gccccatgcg	atgccgggtga	23340
ggatctcgaa	ctcgggtgtcg	caggtgttgc	cgccgatcac	cttcacgcgc	aggcgccggg	23400
tggtgcccc	cgggccgcgtc	gcgcgcgagga	acggcagcg	gtcgcgcgcg	agcgccacgc	23460
cgaacggggc	caggtcggcg	agcgactcca	tcatacaggaa	gatgatgtcg	ggcgctccg	23520
cggtggccgg	gcgctcgcc	gagcgctcct	ggcgctccag	cagccgctcc	acgccctcgc	23580
gcagcccgcg	cgccgcgacc	gggcggggcga	tgctcccggc	gaccagcacg	gcgtgcagga	23640
cgaagccag	gcgcgcgacg	ttcacgttgt	ggtcgggtgg	cgaatacgtc	atcgcgctcga	23700
actccaccgg	ccagcgggcg	aagcgtccgg	cgacgaaggc	gacgccgagc	gcggcgagca	23760
cgagcagcca	ctgccggcgga	tcgtcgtacg	cgtcgtgcag	gtcgatggcc	tcgggtgatct	23820
cgaacgcggc	ccacgcgcagc	gccgcgagca	ctgcggcagc	ggcaagcgcg	cgccagccca	23880
cgcgcgggcc	atgcaggtac	gccggcgcgga	acaccagcgc	ctcgcgagcg	ctcaggaaat	23940
cgccccggcc	gaccacgcgg	ccgaagaaggc	cgagcgcccc	tgcggttcacc	gcgtagagcg	24000
cgcccgccag	cgccacgggtc	cccaccgtcg	cgagcgcccc	gaccgcgacg	aagagcagcg	24060
catggaagat	ggcgagcagg	ccggcgttcc	aggccaggcg	ccgcgcgcgc	gccacgggtt	24120
ccacgtcata	gcccaccacg	aggcgccgct	cgaggcgggc	ctgcagcacc	gcgtaggcga	24180
gcacgcagac	cagcgtgccg	cccagcagca	cgccgaggaa	acccgccaaa	ccagccgtgt	24240
cgtcgaggat	catgggacca	ccagggacgt	cttcgggcca	gtctataacg	ccccggccca	24300
tgaaaaacgg	cgaggccccg	gggcctcgcc	gtggagggaa	ccgcgcaacc	tcagggttgc	24360
gggggtgcga	ccgaccagcc	gccggggcg	atcaggtttg	cgctgccgat	gggtggcgagc	24420
ccgttcatca	ggcgaccgt	gacgctgcgg	tcgggtgtgg	ggaagaccag	gtccttgccg	24480
ctgtcgccgt	tcaggtccag	cagctgcgtc	acgttccaac	ccgtgcccgc	cggcagcgca	24540
tcgcccgcgg	tgatgatgga	aacaccatcc	atcaggcgga	tgtgcacgcg	gccgtcggtg	24600
tggcggaaga	ccaggtcgtc	cttcgcgtcg	ccgttcaggt	cgcccaccag	gttcacccgac	24660
cacccgctgc	cggccggcag	gagctggggc	ctggcgccga	aggtgggtgcc	gttcatcagg	24720
aacaggtgcg	cgcgcccgct	ggtgtggcg	aagatcatgt	ccgcgcgtcc	gtcgccggtt	24780
aggtcgccca	ggtggcttac	cgtccagccg	ctggcgggcg	agaggaagcc	cgctccggcc	24840
gtcacccgtg	tgccgtccat	gatgtagatg	tagccgcggc	cgtcgtgtgt	catgaagacc	24900
aggtccgcct	tgccgtcgcc	gttcattgtc	ccgcctggg	tgagcgtcca	gccgggtccc	24960
gccccgaaga	gctgggcgct	gccgatgatc	gtcgtgccgt	ccatcaacca	gatgtgcgcg	25020
cgtccgtcgg	tgtggcgag	gatgatgtcc	gccttgccgt	cgccgttcag	gtccgcgggtg	25080
tggtcacga	tccagccgag	cccggcgggc	agcagctcct	tgccggccgt	caccgtagtc	25140
ccgttcatga	tgtacgcgta	ggcgcgggca	tcgggtgtgg	ggaacaggat	gtcggcgcg	25200
aggtcgccat	tgaggtcggc	cgtctggctc	accgtccagc	ccgcgcggcg	gccgatcagg	25260
ttggccgagc	cgggtgatggc	ggtgccgttc	atcgtccaca	cccgatgcg	cccgtcgctg	25320
ttctgcagca	cgaggtcgct	cctgcgtcg	ccgctcaggt	ccgagacgac	cggttggtg	25380
ggcgggcgtg	cgccaccgga	aaccggcagt	gccagcccg	cgatccaggc	gcgggtcggcg	25440
cagttggcgc	catccggggc	cgggcggttc	caggtcgcg	agttgcaagg	cgtggagagg	25500
aagttcgaga	aacgccacac	cagcgtgtgc	gcgcccgcgg	tcaccgggaa	cgacaccagc	25560

ttccagcccg	aggccgtggt	gcccgcgtcg	ctgaacacca	ccgtgccgtc	gatcaggaac	25620
tcgaacttgc	cggcgttggg	gaagctcgac	acgcgatagg	cgaacgcgac	gttgcccgcg	25680
agcagcgtgc	ccgcgaacga	gaggtcggag	ttcaccgtcg	tcgagttggt	cgggtcgctc	25740
gtcaccacct	gcgcgcaacg	aaggctggtc	gcgcctcga	aggcctggtc	ggagccgacc	25800
gtccatgcgt	tggcgccgcc	cgaggtcgtg	aagccggcag	gcagcgtgcc	accctggggc	25860
cacggcgaga	ggatgatgcc	ggtcgcggtc	gccggggccg	cgatgaagac	accctgggaa	25920
cccgtcgcat	tggagagcgc	gacgttgaac	gtctcgttgc	cttcaccgat	accgtcgctc	25980
gccaccggga	ccgtgatggt	cttggggcgtg	gtctcgccgt	tggcccagct	gagcgagccc	26040
gaggtcgcgg	tgtagtccgc	gcccgaagtg	gcggtgccgt	tggaggtggc	gtagctgacc	26100
gagatggcac	ccgcgagacc	gccgatgcgg	ctcacgggtga	gcgtcacgtt	gccggcggtt	26160
tccgcggccg	cgaaggtggt	gccggtgaac	tgcaccgaac	cgggaacggc	gggcggggtg	26220
accgtggact	tgagcgccga	gagtgccgcg	cggttgttgt	tcagcgagag	cgctcgcttg	26280
gcctgcgcgg	tgccgcacgg	ttgcggcgca	ccgacaccgg	tcgtgcacga	gaggcccggg	26340
ttggagaagc	ggtagacctt	ggtcgtcgag	tcatgcagca	gcgacatgat	ggtggtgaag	26400
ttcgcccgct	ggttcgctcg	gcactcgggg	cgctggctca	cgccgcaggc	gccgcgcggc	26460
gagccgttgg	gcaggggccga	ggtgcaggtg	agggttgccgc	tcgcgcaata	gaagtggccg	26520
tacgagtacg	cgaacgcacc	ttgcggcggg	ctcgcggtgc	cgcccgcttc	ccaggcctgc	26580
gaggggcggt	cgtggcggtt	gcccattggca	tggccgagct	cgtggatgaa	gaccactcgc	26640
caaccgaaca	ggcagccgga	agcaacggcg	taggcaaagt	tggggtcggg	cgtgctgggt	26700
ccgatccagg	ctgcgcctgt	gccgcggaag	tcggcgccgt	tgccgatgaa	ggccaccatg	26760
tccgcgcctg	attgcgtgcg	gatggcctcg	atgttgccga	acgtcgcggc	gtcgaagctc	26820
gcgtcgcccc	gggtgatggc	atgcatacgc	gtggtgtcgt	cgatggcgct	gtcgtagtct	26880
acctgcgtcg	cattcaccag	gcgcagcgtg	atggcgacct	cgctgtcggc	atacgcagtg	26940
ttggcgcgcg	tgacgaggaa	gttcaggcgc	gtcatcaggc	tgccgcccag	gcgctggccg	27000
aagccgcgcg	agtacacgat	catcaggtcg	atgacgttct	gcggcggtgg	cgtcacctgg	27060
gcgagccgac	ggtattcacg	ccgggggatgg	cgatcagggg	ctcgggggctg	ccggctcgca	27120
cgacgcgcga	cttcgagccg	gtcgtggcca	cggcggtggtc	ggcgccgagg	ctgatgatcg	27180
gcatgctggc	ctgctcgccc	gtcatgtcga	ccaggtagtc	gacggtgccg	cccggaatca	27240
ggcggtactc	accctgcggg	gtcgagaaca	cgccgaagga	gccttcaggc	ccggtcgtga	27300
cgatcgcgcg	gtggttgatg	ccgtcggtct	tgctcttcgc	gatccacgac	gtgatgccgt	27360
cgccgtgggg	ctggagcaac	tcgaagacat	acgagtggcg	caccccgctt	gggagcgaca	27420
gctcgacctc	ggagcgcgga	ctcagggcat	ggagggccgc	ggcggttgaag	cgcaccgctt	27480
gccgggcaac	ggcgctgccc	ggcgccggca	tcgcggcggc	gggtgcgagg	accaggatct	27540
ggggaaaccgc	ggcgaaggcc	acggaagaga	agacgatcgc	cagcgcgccg	aggaagcgcg	27600
ctgtgatccg	ttgcgtgaat	gcgttcatgt	ggagctccgg	aagttgacct	atgcccaatc	27660
cgctatgtcg	cgagatgtg	gacaaaaggt	atcgagcggg	cgtgacgacc	cgcccccgga	27720
gggatgctcc	aaaaggacta	cgagggtgct	acggctgggt	tgggtggcgaa	ggccgaagga	27780
gaactccttg	tggctctccg	gcgacttaag	ggtgcgaggg	aaccaccgac	caccaccgcg	27840
ggccgatcaa	gtccgcgctg	ccgatcgtgg	tgacgcctgt	catcaggcga	atcgtcacgc	27900
gcccattcgt	gtggtggaag	accacgtcca	tgccgccgtc	gccgttgaag	tcgaggagct	27960
gcgtcaccgt	ccagcccgcg	aacggcgcca	ggatgtcggc	ggcgccgagg	atcgcggtgc	28020
cggtcatcag	gcgcacgtgc	gcgcggccgt	cgggtgtcgc	gaacacgagg	tccgcccggc	28080
cgtcgccgtt	gagatctccc	accaggttca	ccgaccatcc	ggtgcccgcg	ggcaacagct	28140
cggcgtggc	gccgaacgtc	gtgccatcca	tcaggaacag	gtgtgcgcgg	ccgtcgttgt	28200
ggcggaagac	gaggtcggcc	ttctggtcgc	cgctgaagtc	cgcgacgtgg	ctcaccgccc	28260
agccgctgcc	gggcgcgagg	aagcccacgc	cggcggtgat	cgccgtgccg	ttcatgatgt	28320
atgcgtagcc	gcgcccacgc	acgttgagga	agacgatgtc	ggccttgccg	tcgcccgttc	28380
tgtcgccggt	gccgaccacg	ttccatcccc	agcccgcggg	cagcaggctc	gcactgcgcg	28440
tgatcgcggt	gccattcatc	agccagatgt	gggcgcgtcc	gtccgtgtgc	tgagcagcag	28500
gtcggccttg	ccctcgccgt	tcaggtccgc	cgtgtggctg	atggtccagc	ccgtgcccgc	28560
ggggaaacag	tccttgccgc	ccacgaccgt	gagcccgttc	atctggtacg	tgtagatgcg	28620
ccgctcggtg	tgctggaagt	agatgtcggc	ctggctgtcg	ccgttggaat	ccgcggccat	28680
tgatcacgct	ccagccggcg	ccggcgggaa	tcaggttcgc	cgaaccagtg	atcgtcgctc	28740
cgttcatggt	ccacgcggcg	atgcgcccac	cgggtgtgctg	gaacagggaag	tcgctctgcc	28800
cgtcgccgtt	caggtcgggtc	gccgcattgg	tcgggaacgc	caccttcgcg	aggaaggcat	28860
cggagaaacc	cgtgagcaac	gtcttgtacg	cgccgggcgt	ggtcggatag	ttcgccgacg	28920
acgtgcggcc	ggccacatag	gccgcgcccc	gogcatcgac	gccgacgcgg	tacccacact	28980

cgacgttggt	gccgcccgatc	agcgtcgaga	acacgatgtc	gccgttgccg	acgcgagtga	29040
ggaacgcgtc	ggcatcgccg	ccggggccct	gcaccgggtt	cagggtcggg	aagccggcga	29100
cctgcgtgta	gcccacgacg	agcgcacgc	cattcgccgc	cagggcgacg	tccaggggat	29160
tgtcgtaggc	cgagccgccg	aagatgcccg	acgattcgag	cgcccggtg	gccgagaagc	29220
gcgtcacgaa	cgcacgacg	gtgccggcga	aattgcgcac	cgttccggtc	tgccgggaagc	29280
tgggcaggcc	ggtatcgccg	acgacggtcg	cctggcccgt	gaccgggttc	accgcgatcg	29340
cggtcacgtt	gtcgccgctg	ccgccgagga	acgtggaata	cgcgaccggg	ccgccgctcg	29400
caccgatctt	ggtgacgaag	ccgccaccgg	ggtgggtctg	gtacgcattc	acgttgcgga	29460
acgtcgagga	gcgcgtggag	cccgcgacgt	acgtgttgcc	cgcattgtcc	gttgcgatac	29520
cgccggccgat	gtcatcaccg	ctgccgccga	gatacgtgga	atagacgaag	ccgccgttga	29580
acgagagctg	gtgacgaacg	cgtcgcgatt	gaggccgggtg	ccgacgaaca	cggtcgagc	29640
gcgctgccgc	cggtcgggaa	gccgcgcgcg	gccgaaccgg	tgacatgcat	gaagccgttg	29700
ttgtccacgg	cgatcgccat	gccctcgteg	tcgccggagc	cgcccagcag	ccgcgagaag	29760
atgacgttgc	cagtggattt	cagcgccgtg	acgaatgcat	cgccccgcc	ggcatggttg	29820
gtgcaatcgg	tgacggcgaa	gcaatgcggg	aacgcgttgg	tgccggtgga	accggtgatc	29880
gcgacgaact	cgttgccgcc	ggaggcgccg	agcgccagcc	cgcgagata	cgcttcggcc	29940
ccgagcgtgg	tcgagtacac	cacctgctgc	gtgggtgggt	tgaacttcac	caccgtggct	30000
tcgtacgtgc	ctgcggtata	ggtcgcgagc	ccgacgtaca	cgttgccctg	cgcgccacc	30060
ttgaccggcg	tggtccactg	ctcgctcgccg	gcgccgccca	ggtacgtgga	gaacgccatc	30120
accgggtcga	tcaccagcgt	gcgggtgctg	tcgtagtcgg	cgatcacgaa	gccggccctcg	30180
gcgccttgct	cgcccacgaa	gagctcgaag	cgccggccca	ccggcacgcg	cgaatcgccg	30240
acctgctgga	aggctaccgg	cgcgctgctg	gtaaattcct	cctggcccac	gcgcatgcgc	30300
aggttgccct	cgccatcgat	ccacgcgtgg	tcggcggacg	agaggtcgag	gcggatctgg	30360
cgcggtatcgg	cgcgcggggc	gaccacgaag	tcgtattcga	gcgtgccctc	cttgccgtac	30420
acgggtgaggt	ccacgcgcgc	atacaggtcc	ttcagcgaca	cgcgcccgaa	gtgcggcacg	30480
ttttcgcgcc	ggccggcgcc	cgtggtgccc	ctgtaatagt	ggctgagggg	ggcccggggc	30540
tcctcgccct	cgatgaccgg	atcggaagcc	cccgcgaagc	gcacgcgcaa	cagcgagcgc	30600
ctggcagctt	tcctggagcg	tggctcgaag	gcaatgccat	cgcggtcac	ggagacgcgt	30660
ccggcctgcc	cgcgcgagac	gtagagcgca	tcggggccga	actggccctc	gttgcgctcg	30720
aacgtgatgg	gcgcgttcct	gagggcatca	gggacgacgg	catggacggc	ggcgggagtg	30780
gcgagggcgg	ccgcgacgag	cacggccagg	ggcgggagcg	ccgggcccgg	gccgctggag	30840
ctgcgggtca	tggggcagtg	acctttctgt	tggtatggct	gcttggttgc	tttgcgggat	30900
ccacagctcg	gatccagctg	gcgatactac	cagcgagcga	gtggtttttc	gtgaccagtg	30960
cccgcagctg	ggccccggcg	acaacggcgg	catcgcggtc	ctttgccagg	gcacgcagcg	31020
cctcgatcca	gtcgccgtcg	cgggccagca	cggcgggggc	atcgcggtag	ggctcgagat	31080
ccgagggccac	caccgggatg	ccgagggcgc	cgctactcgag	cagcttcagg	ttgctcttcg	31140
cgcgattgaa	cggggttgctg	cgcagcggtg	ccaccgcgac	gtcgagtgcc	agcgacgcga	31200
gctttctcgg	gtactgcgcg	atgggcacca	tgctcgtcac	ttccgcggcg	aacggcgcca	31260
gctcgggcgt	gcacaggccg	aggaacaccc	agtcgatctc	gcgatgcgtg	gcgcgcacca	31320
cgggctcgag	caacttcagg	tcttccccat	gctgtttcgc	ccccgccag	cccacgcgcg	31380
gacgcgcgcc	acccgcggga	cggttcgcga	ggccggccca	acgtccgca	tcgatcgctg	31440
tcgggatcac	gcgcacgtcc	ttcgcgccgc	ggccgaaggc	ctccgcgagc	ggtgcggtgg	31500
acaccacgag	gcggtcgcac	aacgccaccg	cgcgcgcgat	gcgctggggc	atgtccgggt	31560
agatcgtcgc	cgcatacgga	ttgccggggc	gcagttgcgt	gagcaggtca	tccaggccca	31620
ggaccttcag	cgcgcgctcc	tggcgcgcgga	gcacctcgag	cgaggtgagc	tggtagtgtg	31680
ggaagaagtt	gtgcgcgagc	acggcgctcg	cgtcgaggcg	ctgccactcg	acgcggttgg	31740
gcgcgcagcc	cgtggcgctg	tcgcccatca	tcacgacctc	gacgcggccc	gcgcgctcca	31800
gcgcgcgcga	cggtggcgcg	acgcgcacct	cgccggagcc	ccagcggtcg	aacggaaacg	31860
cgcagagccg	aaccgcgtgc	gtgtcgttgc	cgcgcgcgcg	gaagcgttcc	acgggcccgt	31920
agccctcgcc	gcccatgcgc	aacgccgggt	ggtagtgcgg	atcgctcgctg	agcaccggct	31980
gccagcgctc	gcgcacccac	gcctcttcgg	aaggaatgac	gagcttcgcc	accggaacga	32040
cggtatcgaa	gcccgcgtcg	cgcagccccga	ggaagtccgc	cgcgatgcgc	gggcgcgcaa	32100
cgcgctcgat	ctcgctgcagc	gcctgcgcgg	agagcgccgc	gtcgcggttg	accagcgcgga	32160
gacggccccg	cgcgacgctc	acctcgcgcg	gcgatgcgta	gagtgcgcgc	aggcgctcct	32220
cgccctgcgc	gcgcggcgga	gcgcccgaga	cggcccacgg	cccgcgcgcc	gcgatttccc	32280
agccgggaac	gcgcacaccg	ccggggccca	gcaggtcggg	cgcgagggca	cccgtggagg	32340
						32400

ggcctgcat	gcectgctcg	aggcgctcga	gccagccggc	gcggaattcg	ctgcaacgcg	32460
catcgacgat	cgcatccac	agcgtcgccg	cgttgccgat	ggcgcgcgcg	aacgcggagc	32520
cggtttcgcc	gcccaccgga	accaggtgcg	tgcgcacgcc	gcgcccgcgc	agctcaccgg	32580
gccagcgttt	cggcgccctt	cgcgctcgaga	tgcagctaca	cgtggccgat	acgttcgcgg	32640
cagcccttca	cgagcgccct	gatgcacgcc	accagccgct	ccggcgcgctg	cgctcgctca	32700
atgaaggcgc	aaatcgtgcc	ggccccgatc	gcgcgcgccc	acgatgcggg	cgcatcggac	32760
gcagctggca	ggaacgcggc	ctgcgtgcgc	gtgcggcgga	agtgtctcgc	gaccacgcgc	32820
tgggtggcgct	cgcgcgcgcg	gtcgctttcc	tgcgcggcga	tgttcaggct	gaggcgatgc	32880
acgagcagct	cgcgagtgct	caccacggcg	ctcgacggcg	cctcctccgc	gaggcgagc	32940
gcgagatcgg	tgacgcccgc	ccagccgcgc	gcgcgagcc	cgccacgctc	cagcaacgca	33000
gcgcgggcga	gcaccgcgag	cccggcgagc	ggctsggtgtc	gcggaactc	cgcgccgaac	33060
gtggggacga	acgcccggatc	gtagcgcttcg	ccgtgcgcgc	tcagggcac	actgcagccg	33120
tagagaagct	cggcggttcgg	gttgcggtgcg	accgcaccgc	ccaccgcgct	cgcgcgccca	33180
tcggccagcc	ggctgccttc	ctccaccagc	acgaccatt	cgcccttcgc	gtccgcgctc	33240
gcgcggttga	cgcgcccagc	cgccgcgctc	ctcgcgctgc	gctcgaggac	cacggtctcc	33300
agggcgatcg	cggcagtgct	cctggcgggc	aaggcctcgc	gcgccaacca	gtggtcgtag	33360
gcatcgccgc	ccagcgcgct	gcgctcgcg	cgcgctggaa	agccgaggca	cgccgcggcg	33420
aacaacggat	cgatcgttgc	gggcacagca	tcaccgcgc	catcgcgctc	cgccgagcgc	33480
tcgcataga	gcgccaggta	ctcgcccgca	ttcgccctga	tcgtcttcgg	gcggtggtac	33540
ttcaggttct	ccgcgacgcg	atcgagcgcc	gtgcggctgt	cgcgagcgcg	cagcaggagg	33600
gccacgatcg	cgccgggac	gttggtgcgc	agcaggagc	cggtctcgcc	atcgcgacg	33660
cgctcgggaa	tgccgcccac	gcgctgccc	gccaccggga	tgccgaggcg	ctgtgcctcg	33720
ctcaacacca	ggcagaagg	ctcctccac	gtcgagggca	gcaggaccag	gtcgatgccg	33780
cgaaggatgc	gcggcaggct	cgctcgcgca	taggcgccc	gctgcagat	gcccgcgcga	33840
ttcgccgcgc	cttcgagcgc	gagatcgcg	ggacccagc	cctcgatcgc	gatgcgctgc	33900
cccgcagcc	ggcgcgcggc	ctcgatgatc	aggtgcgcgc	ccttgccccc	gccgaagcgg	33960
ccgaggaagg	ccacgcgcag	caccggccgc	gcttcggggc	gcggaagcag	cggcaggctcg	34020
gcaatgccgt	gcccgatgac	gcgcagcttc	gcttcgcagc	tcgcgcgaa	gcccctgcgc	34080
atgcgcacgc	ccgcgtactc	ggacggcgca	aggatcacgt	cgccgcgcgc	cacggcatcg	34140
gacgcgcgcg	cgaagcgca	ttcgaggtag	ccggtcacgg	acgcggcgcg	cgctcgcccc	34200
acgatcgct	cggtcgggcc	gcgcaggcaa	tcacgcac	cgccatcggc	gcccgcgcgcg	34260
gcggcgcgcc	cgcacggcag	gtcctgcggc	ccggtcatca	tgttgtagtc	ggcgcacagg	34320
tacgagaagt	cgtgcgcgcg	gatgaccacg	cgtgcccccg	actcacgcgc	gatgcgcggc	34380
aggcgagggc	tgttcacagc	caccaacgac	tggaagtgc	cgacgtcgta	gcccgcggac	34440
agcagttcgc	ggaaaggccgc	ctccacgccc	gcgtcgccaa	ccgcgcctc	gtgcccgatc	34500
acgcggatgc	cgggcgctcat	gcgctcgggg	tcaggcgcg	cgagtcgcac	gtgcgtcgcc	34560
aggcgctcgg	tgttgaagtc	ggctccacgcg	cgcgggccct	cgggcagcac	cagcgtcgag	34620
gcgacgtcct	tgcgacgag	gttcaccagc	gcgcgcggtg	gctcctcaat	gcccgcgcgc	34680
gagtcgaggc	gatgcagcac	ctgcagcacg	cgcgggcggc	ccggcatccg	ctcgcgctcg	34740
gcgcggcgct	tgacgcgctc	ggcgggcgcg	cgagcggggt	tcgcgatcgc	ccacgcgcgc	34800
acgcccgcga	cgtaggtggg	ccagcgggcg	tcgaggcgca	cgcggttggc	ctcgcgcgcg	34860
gcgtcgatgc	cggccacccc	gccgaacgag	gcctcgcccg	cgtggtgcac	gtagacgtcg	34920
tcgcagcacg	cgatctgcca	gtcccgggtc	agcgcgcgca	tcgagaggtc	gttttctctg	34980
ccgtagccgc	ggccatagag	cgagtcgaac	gggcccgcatt	gcacccacac	ctcgcgcgcg	35040
accagcatgc	agaagccgac	cgccgtcggc	aggcgcggt	actcgcgcg	gctggcgcg	35100
gcgacgcagg	cggagatggc	atcgacgtcc	ggctcgccgt	cgggccgctc	gaagatatcg	35160
cgacgcgcgg	gcaggctgag	gatcgctcg	cggttgagga	gcgggcacac	gatgccgacc	35220
cgcgcatcgc	tgtcgcgcca	acgcagcagg	ccctcgagcc	aaccggggcc	gacctgcgta	35280
tcggaattga	gcaccacgaa	gtcgccgccc	ctgcggcgca	ggcccgcagc	gaccgcgct	35340
gcgaagccga	ggttcttcgc	attggtgacg	acacgagtcg	cggcacgacg	cgccgcgaag	35400
ctgcccagca	acccggcgag	gcgcggatcg	gtgctcgctg	catcgacgag	gacgatgcgc	35460
tgcgcggcg	cgtgtggcg	ggccagcgac	gcgaggcagc	gttcggtgtc	gtcgtggccg	35520
ttgaagaccg	ggacgacgac	tagcgccggc	gccatccgac	cggcgcttcg	agctggggcca	35580
atcgcgccag	caggcgctgg	tgctcctgcg	agagtgccct	gagctgcttc	tgcgaggcga	35640
cgtaggcgcc	tccatgcgct	cccaggtgcg	tgcgcgccct	ctcgagctgc	cgacgcgact	35700
gctcgtggcg	accggagacg	tcgtgggtcg	ggctcgagag	cagcgagaag	cgcccggccc	35760
ccgcccgtcac	cgccctcgcg	tcgcccgcaca	gggcgaggaa	gtacatcgca	tcggcgagcgc	35820

ccgccgcggg	tttcgcagcc	tccggcgccg	aagcttgacg	caactgcacg	gcgtcgggca	35880
ccgcgtgcag	eggccagatc	gccgaatacg	catcgacgcg	ttgcccgaac	agggccacct	35940
ccgggaaccg	cgccttcagc	acggcgagga	attcgtgttc	gtagagctcg	cgcacgtgat	36000
gcgggttgcg	gtagtgcgcg	tggtcgaggt	acacctcgcg	gttcggcgctc	gagaccagca	36060
gcaggccgcc	ggggcgccagc	acgcgcttcg	cttcgtcgag	caggcgctcg	ggatcgggaa	36120
tgtgctcgag	cgtctcgaag	gagacgagga	gatcgatgct	tgcgtcatcg	cacggcagcg	36180
actcgacgcg	gccctcgacg	tactgcaggt	tctgcgcggc	accatagcgg	cggcgcgctt	36240
gcgcgatggt	ctcggcgggc	acatccgcgc	ccaccacgga	cttcgcgcgc	gtcgccagca	36300
gcgccgagcc	gtagccctcc	ccgcaggcga	tgtcgaggac	gcggcagcct	ccggcgagcg	36360
gcaacgcaaa	gtggtagcgg	tgccagtgtc	cgtaccagat	ctcgccctcg	aagccgggct	36420
ggaaacgttc	gtgttccatc	gggagagtat	aggtggatgt	gcaaccgccc	tccggagagg	36480
gcggttgcg	tgatcttatt	cgagatcggc	tgctacggaa	gcaccggcaa	cggctgcggc	36540
ggcgccaccg	accatccggt	gcccgcgccc	agcaggttcg	cgctgccgag	cgtggtgagg	36600
ccgtccatca	ggcgcacctg	gatgcggcca	tgcacattct	tgaacaccat	gtcgagcttg	36660
ccgtcgccat	cgaagtcgag	cagttgcgtg	acggtccagc	ccgcccccg	cggcagcacg	36720
tccgcccgcg	tgaggatggc	ggtgcccgtc	atcaggcgca	cgtgcgcacg	gccgtcgggtg	36780
tggcggaaca	cgatgtcacc	gcgcccgtcg	cggttcatgt	cgcccacgtg	gctcacgac	36840
cagccggtgc	ccgcgcccag	gagctcggaa	ccggcgccga	acgcggtgcc	gttcatgatg	36900
aagagatgcy	cgcggccatc	ggtgtggcgg	aagaacatgt	ccgccttgcc	gtcgccgctc	36960
acgtcgccca	ggtgcgtcac	cgtccagccg	ctggcgggcg	agaggaagcc	cgcgccggcg	37020
gtgatggtgg	tgccgttcac	caggtagatg	tagccgcggc	cgtcagcgtg	gatgaagacg	37080
atgtcgtccc	tgcgctcgcc	gttgaggtcc	ccggtggcca	cgactctcca	gcccgtcgcc	37140
ggccccagca	gctgctggct	gccgatgatg	gccgtgccat	ccatcagcca	caggtgcgag	37200
cggccgtcgg	cgttgcgag	caggagatcc	gccttgccgt	cgccgttcac	gtcggcgggtg	37260
cggtcgatgc	tccagccgag	gccggcgccc	agcagctcct	tgcgcgggt	caccgtgagg	37320
ccgttcacgc	tgtacacgta	cacgcgcccg	tccgtgtggt	gggatcctct	agagtcgacc	37380
tgcaggcatg	caagcttgag	tattctatag	tctcacctaa	atagcttggc	gtaatcatgg	37440
tcatagctgt	ttcctgtgtg	aaattgttat	ccgctcacaa	ttcacacaac	atacgagccg	37500
gaagcat						37507

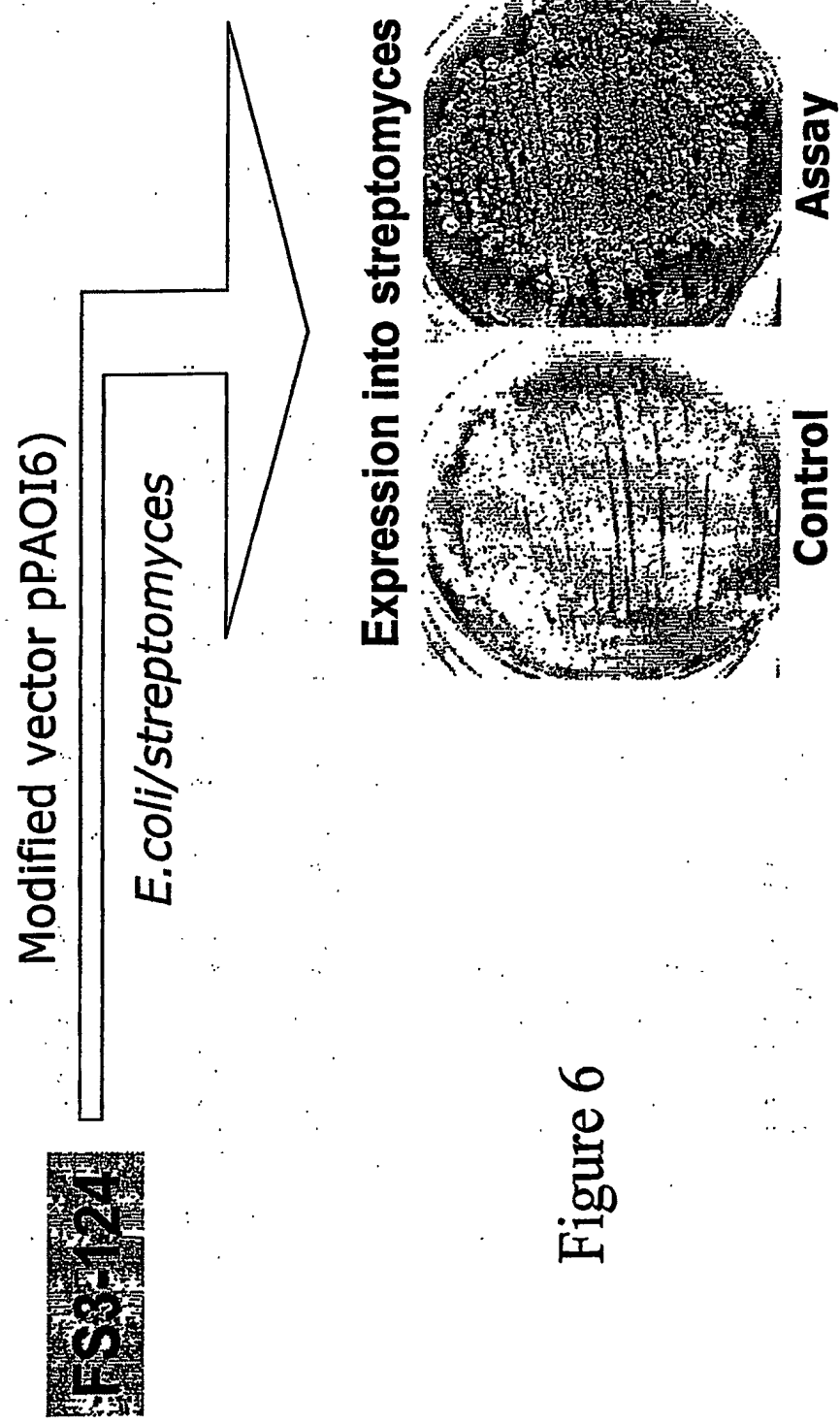


Figure 6

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.